

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
ESCOLA DE ENGENHARIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO**

**ÉVERTON MIGUEL DA SILVA LORETO**

**SISTEMÁTICA PARA SELEÇÃO DE VARIÁVEIS E  
DETERMINAÇÃO DA CONDIÇÃO ÓTIMA DE  
OPERAÇÃO EM PROCESSOS CONTÍNUOS  
MULTIVARIADOS EM MÚLTIPLOS ESTÁGIOS**

**PORTO ALEGRE  
2014**

ÉVERTON MIGUEL DA SILVA LORETO

**SISTEMÁTICA PARA SELEÇÃO DE VARIÁVEIS E DETERMINAÇÃO DA  
CONDIÇÃO ÓTIMA DE OPERAÇÃO EM PROCESSOS CONTÍNUOS  
MULTIVARIADOS EM MÚLTIPLOS ESTÁGIOS**

Tese submetida ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal do Rio Grande do Sul como requisito parcial à obtenção do título de Doutor em Engenharia, na área de concentração em Sistemas de Produção.

Orientador: Michel José Anzanello, PhD.

Porto Alegre  
2014

ÉVERTON MIGUEL DA SILVA LORETO

**SISTEMÁTICA PARA SELEÇÃO DE VARIÁVEIS E DETERMINAÇÃO DA  
CONDIÇÃO ÓTIMA DE OPERAÇÃO EM PROCESSOS CONTÍNUOS  
MULTIVARIADOS EM MÚLTIPLOS ESTÁGIOS**

Esta tese foi julgada adequada para a obtenção do título de Doutor em Engenharia e aprovada em sua forma final pelo Orientador e pela Banca Examinadora designada pelo Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal do Rio Grande do Sul.

---

Prof. Michel José Anzanello, PhD.  
Orientador PPGEP/UFRGS

---

Prof. José Luis Duarte Ribeiro, PhD.  
Coordenador PPGEP/UFRGS

BANCA EXAMINADORA:

Prof. Cláudio Alcides Jacoski, Dr. (UNOCHAPECÓ)

Prof. Flávio Sanson Fogliatto, PhD. (PPGEP/UFRGS)

Prof. Marcelo Farenzena, Dr. (PPGEQ/UFRGS)

*Dedico esta tese à **Mara** e ao **Eduardo**, esposa e filho,  
que sempre me apoiaram nesta longa caminhada.*

## AGRADECIMENTOS

Agradeço ao meu orientador, Prof. Dr. Michel José Anzanello, pela acolhida e suporte no momento mais difícil desta caminhada, pelo seu companheirismo e profissionalismo, que possibilitaram a conclusão desta tese.

Agradeço à minha esposa Mara e ao filho Eduardo, pelo estímulo e compreensão durante todo este período.

Agradeço aos meus pais, Miguel de Loreto (*in memoriam*) e Eulina da Silva Loreto, pelos valores que me ensinaram e pelo amor que sempre me dedicaram.

Agradeço à UNOCHAPECÓ, na pessoa do professor Dr. Cláudio Alcides Jacoski, pela oportunidade criada através deste Doutorado Interinstitucional em Engenharia de Produção em convênio com a UFRGS.

Agradeço aos professores do Programa de Pós-Graduação de Engenharia de Produção (PPGEP/UFRGS) pelos ensinamentos e sugestões recebidas.

Agradeço aos colegas de doutorado pela amizade, apoio, críticas e sugestões.

A todos que em algum momento, iluminaram o meu caminho, muito obrigado.

E a Deus, por tudo.

## RESUMO

Esta tese apresenta uma sistemática para seleção de variáveis de processo e determinação da condição ótima de operação em processos contínuos multivariados e em múltiplos estágios. O método proposto é composto por seis etapas. Um pré-tratamento nos dados é realizado após a identificação das variáveis de processo e do estabelecimento dos estágios de produção, onde são descartadas observações com valores espúrios e dados remanescentes são padronizados. Em seguida, cada estágio é modelado através de uma regressão *Partial Least Squares* (PLS) que associa a variável dependente daquele estágio às variáveis independentes de todos os estágios anteriores. A posterior seleção de variáveis independentes apoia-se nos coeficientes da regressão PLS; a cada interação, a variável com menor coeficiente de regressão é removida e um novo modelo PLS é gerado. O erro de predição é então avaliado e uma nova eliminação é promovida até que o número de variáveis remanescentes seja igual ao número de variáveis latentes (condição limite para geração de novos modelos PLS). O conjunto com menor erro determina as variáveis de processo mais relevantes para cada modelo. O conjunto de modelos PLS constituído pelas variáveis selecionadas é então integrado a uma programação quadrática para definição das condições de operação que minimizem o desvio entre os valores preditos e nominais das variáveis de resposta. A sistemática proposta foi validada através de dois exemplos numéricos. O primeiro utilizou dados de uma empresa do setor avícola, enquanto que o segundo apoiou-se em dados simulados.

**Palavras-chave:** PLS, Seleção de variáveis, Processos Multivariados; Processos em Múltiplos Estágios.

## ABSTRACT

This dissertation proposes a novel approach for process variable selection and determination of the optimal operating condition in multiple stages, multivariate continuous processes. The proposed framework relies on six steps. First, a pre-treatment of the data is carried out followed by the definition of production stages and removal of outliers. Next, each stage is modeled by a Partial Least Squares regression (PLS) which associates the dependent variable of each stage to all independent variables from previous stages. Independent variables are then iteratively selected based on PLS regression coefficients as follows: the variable with the lowest regression coefficient is removed and a new PLS model is generated. The prediction error is then evaluated and a new elimination is promoted until the number of remaining variables is equal to the number of latent variables (boundary condition for the generation of new PLS models). The subset of independent variables yielding the lowest predictive in each PLS model error is chosen. The set of PLS models consisting of the selected variables is then integrated to a quadratic programming aimed at defining the optimal operating conditions that minimize the deviation between the predicted and nominal values of response variables. The proposed approach was validated through two numerical examples. The first was applied to data from a poultry company, while the second used simulated data.

**Keywords:** PLS, Variable Selection, Multivariate Process, Multistage Process

## LISTA DE APÊNDICES

Apêndice A - Matrizes de correlação por estágio .....	101
Apêndice B - Rotinas utilizadas no Matlab .....	104
Apêndice C - Limites de Especificação, de Controle e Índices de Capacidade do Processo .....	109
Apêndice D – Modelo de Programação Quadrática .....	110
Apêndice E – Matriz de correlação do estudo simulado, por estágio .....	111
Apêndice F – Coeficientes de regressão PLS do estudo simulado .....	117
Apêndice G – Limites de Especificação, de Controle e Índices de Capacidade do Processo do estudo simulado .....	119
Apêndice H – Modelo de programação quadrática do estudo simulado .....	120

## LISTA DE FIGURAS

Figura 1 – Fluxograma dos segmentos produtivos da cadeia principal de aves.....	22
Figura 2 - Fluxograma do processamento de aves .....	24
Figura 3 - Fluxograma do processo de produtos empanados .....	26
Figura 4- Exemplo de um processo em cascata multiestágio.....	28
Figura 5 - Exemplo de fluxos de processamento em processos multiestágios.....	28
Figura 6 - Perfil hipotético na determinação do número de variáveis latentes.....	36
Figura 7 - Limites de especificação de uma característica de qualidade de um produto .....	46
Figura 8 - Função perda de qualidade de Taguchi.....	47
Figura 9 – Perfil do erro de previsão em função do número de variáveis retidas .....	52
Figura 10 – Análise das variáveis quanto à presença de valores espúrios .....	58
Figura 11 – Representação esquemática das variáveis do estágio 1.....	61
Figura 12 – Representação esquemática das variáveis do estágio 2.....	61
Figura 13 – Representação esquemática das variáveis do estágio 3.....	62
Figura 14 – Determinação do número de componentes para o estágio 1 .....	63
Figura 15 – Comportamento do RMSE em função do número de variáveis latentes para o grupo D do estágio 1 .....	64
Figura 16 – Determinação do número de componentes para o estágio 2 .....	64
Figura 17 – Determinação do número de componentes para o estágio 3 .....	65
Figura 18 – Comportamento do $RMSE_p$ em função do número de variáveis retidas no estágio 1 .....	66
Figura 19 – Comportamento do $RMSE_p$ em função do número de variáveis retidas no estágio 2 .....	67
Figura 20 – Comportamento do $RMSE_p$ em função do número de variáveis retidas no estágio 3 .....	68
Figura 21 – Tela de entrada da ferramenta <i>Optimization Tool</i> do Matlab .....	72
Figura 22 – Determinação do número de componentes para o estágio 1 para o estudo simulado .....	79
Figura 23 – Determinação do número de componentes para o estágio 2 para o estudo simulado .....	80
Figura 24 – Determinação do número de componentes para o estágio 3 para o estudo simulado .....	80

Figura 25 – RMSEp em função do número de variáveis retidas no estágio 1 para o estudo simulado.....	81
Figura 26 – RMSEp em função do número de variáveis retidas no estágio 2 para o estudo simulado.....	83
Figura 27 – RMSEp em função do número de variáveis retidas no estágio 3 para o estudo simulado.....	85

**LISTA DE QUADROS**

Quadro 1 - Limites de especificação, valor nominal e índice de importância.....	54
Quadro 2 - Classificação das variáveis por tipo e estágio .....	59
Quadro 3 – Interpretação do coeficiente de correlação .....	60
Quadro 4 - Formação dos grupos para a validação cruzada .....	62

## LISTA DE TABELAS

Tabela 1 – Percentual de variáveis retidas por estágio .....	68
Tabela 2 – Coeficientes de regressão PLS do estágio 1 .....	69
Tabela 3 – Coeficientes de regressão PLS do estágio 2 .....	69
Tabela 4 – Coeficientes de regressão PLS do estágio 3 .....	69
Tabela 5 – Limites de especificação das variáveis de resposta e índice de importância.....	70
Tabela 6 – Limites de operação, média e desvio padrão das variáveis de independentes retidas .....	71
Tabela 7 – Valores esperados e nominais para variáveis de resposta com o modelo proposto	73
Tabela 8 - Condição ótima de trabalho para o processo, gerada pelo modelo proposto .....	73
Tabela 9 – Comparativo de resultados entre os modelos completo e proposto.....	74
Tabela 10 – Comparativo dos valores esperados para as variáveis de resposta pelo modelo completo e proposto.....	75
Tabela 11 - Condição ótima de trabalho para o processo nos modelos completo e proposto ..	76
Tabela 12– Relacionamentos gerados entre variáveis de resposta e independentes .....	78
Tabela 13 - Matriz dos coeficientes de correlação de algumas variáveis do estágio 1 .....	82
Tabela 14 – Percentual de variáveis retidas por estágio, no estudo simulado.....	85
Tabela 15– Condição ótima de trabalho para o estudo simulado .....	87
Tabela 16– Valores esperados nas variáveis de resposta na condição ótima para o estudo simulado.....	88
Tabela 17 – Comparativo de resultados no estudo simulado entre o modelo completo e o modelo proposto .....	88
Tabela 18– Valores esperados nas variáveis de resposta na condição ótima para o estudo simulado.....	89
Tabela 19– Condição ótima de trabalho com o modelo completo para o estudo simulado .....	90

## LISTA DE SIGLAS

API – Active Pharmaceutical Ingredient  
BVE - Backward Variable Elimination  
BQ - Backward- $Q^2_{cum}$   
CovProc - Covariance Procedure  
GA - Genetic Algorithm  
INLR - Implicit Non-Linear Latent Variable Regression  
IPLS – Interval PLS  
IPW - Iterative Predictor Weighting  
IVS - Interactive Variable Selection  
MBPLS - Multi-block Partial Least Squares  
MPCA - Multi-way Principal Component Analysis  
MPLS – Multi-way Partial Least Squares  
MRVL – Modelo de Regressão por Variáveis Latentes  
NIPALS - Non-linear Iterative Partial Least Squares  
PCA - Principal Component Analysis  
PLS - Partial Least Squares  
PPLS - Powered PLS  
PRESS - Predictive Residual Error Sum of Squares  
RLM - Regressão Linear Múltipla  
RMSE – Root Mean Square Error  
RMSE<sub>p</sub> – Root Mean Square Error Pooled  
SPLS– Sparse PLS  
UVE - Uninformative Variable Elimination  
VIP - Variable Importance in the Projection

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>14</b>
<b>1.1</b>	<b>TEMA .....</b>	<b>16</b>
<b>1.2</b>	<b>OBJETIVOS.....</b>	<b>16</b>
<b>1.3</b>	<b>JUSTIFICATIVA .....</b>	<b>17</b>
<b>1.4</b>	<b>DELINEAMENTO DO ESTUDO.....</b>	<b>18</b>
<b>1.4.1</b>	<b>Método de Pesquisa.....</b>	<b>18</b>
<b>1.4.2</b>	<b>Método de Trabalho.....</b>	<b>18</b>
<b>1.5</b>	<b>DELIMITAÇÕES.....</b>	<b>19</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO.....</b>	<b>20</b>
<b>2.1</b>	<b>SETOR AVÍCOLA .....</b>	<b>20</b>
<b>2.2</b>	<b>PROCESSOS EM MÚLTIPLOS ESTÁGIOS.....</b>	<b>27</b>
<b>2.3</b>	<b>UTILIZAÇÃO DE MÉTODOS DE PROJEÇÃO EM CONTROLE DE QUALIDADE ..</b>	<b>29</b>
<b>2.4</b>	<b>REGRESSÃO PLS.....</b>	<b>32</b>
<b>2.5</b>	<b>SELEÇÃO DE VARIÁVEIS EM REGRESSÃO PLS .....</b>	<b>36</b>
<b>2.6</b>	<b>PROGRAMAÇÃO QUADRÁTICA .....</b>	<b>42</b>
<b>2.7</b>	<b>CAPACIDADE DO PROCESSO .....</b>	<b>45</b>
<b>2.8</b>	<b>FUNÇÃO PERDA DE QUALIDADE DE TAGUCHI .....</b>	<b>47</b>
<b>3</b>	<b>METODOLOGIA .....</b>	<b>48</b>
<b>3.1</b>	<b>COLETA E PRÉ-TRATAMENTO DOS DADOS .....</b>	<b>48</b>
<b>3.2</b>	<b>GERAÇÃO DE UM MODELO DE REGRESSÃO PLS PARA CADA ESTÁGIO.....</b>	<b>50</b>
<b>3.3</b>	<b>SELEÇÃO DAS VARIÁVEIS INDEPENDENTES MAIS RELEVANTES PARA CADA REGRESSÃO PLS EM CADA ESTÁGIO DO PROCESSO .....</b>	<b>51</b>
<b>3.4</b>	<b>FORMULAÇÃO DO MODELO DE PROGRAMAÇÃO QUADRÁTICA .....</b>	<b>53</b>
<b>4</b>	<b>EXEMPLOS NUMÉRICOS .....</b>	<b>57</b>
<b>4.1</b>	<b>ESTUDO DE CASO EM UMA EMPRESA AVÍCOLA .....</b>	<b>57</b>
<b>4.2</b>	<b>EXPERIMENTO SIMULADO .....</b>	<b>76</b>
<b>5</b>	<b>CONCLUSÕES.....</b>	<b>91</b>
<b>6</b>	<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>94</b>

## 1 INTRODUÇÃO

A busca pela melhoria da qualidade de produtos gera a necessidade de melhor gerenciamento do processo produtivo nas empresas. A maioria dos produtos se torna mais sofisticada no decorrer do tempo, com a inclusão de novos componentes e materiais, o que também gera processos mais complexos e longos. Aliado a isso, as exigências dos consumidores também vêm aumentando ao longo do tempo, demandando processos mais estáveis a fim de garantir consistência nas especificações do produto.

De acordo com Mandroli *et al.* (2006), os processos podem ser classificados com base no fluxo dos itens: (i) sistema em série, quando o processamento é realizado de forma sequencial em diversas estações de trabalho; (ii) sistema não serial, quando o material de entrada utiliza um dos vários caminhos dentro do sistema de produção, e (iii) sistema de montagem convergente, quando cada estação de trabalho tem, no máximo, um sucessor (mas pode ter várias estações antecessoras). Paralelamente, os sistemas produtivos podem ser classificados em contínuos ou em lotes. As principais características dos processos contínuos são alta velocidade de produção, baixa flexibilidade, número limitado de produtos e poucas interrupções no processamento. Já os processos em bateladas (ou em lotes) tipicamente caracterizam-se por ciclos curtos de produção, produtos mais complexos (customizados) e maior número de etapas de produção (BORGES e DALCOL, 2002).

Um processo serial longo pode ser dividido em estágios, o que facilita seu gerenciamento (KOURTI, 2005). Um estágio pode ser definido por características físicas do processo (por exemplo, quando um produto em processamento passa por um equipamento ou é executada uma operação unitária neste produto) ou por mudança nas características do produto. Os estágios podem ocorrer em unidades industriais fisicamente próximas (anexas) ou distantes, inclusive em cidades diferentes. Segundo Hawkins (1993), processo em múltiplos estágios, ou “processo em cascata”, é uma sequência de produção com valor adicionado em série, onde uma mudança ocorrida em uma variável em determinada etapa só pode ser sentida nesta etapa e nas etapas seguintes a esta mudança (não sendo observada nas etapas anteriores).

Conforme a complexidade de um processo, o número de variáveis a serem monitoradas tende a crescer substancialmente; em processos de múltiplos estágios, esse número de variáveis pode chegar a milhares (CHEN e TSAI, 2009; KETTANEH *et al.*, 2005). Tais variáveis podem ser separadas em: (i) variáveis de processo, que são as medidas de

variáveis controláveis sobre algum parâmetro do processo (também conhecidas como variáveis de entrada ou independentes), e (ii) variáveis de resposta, que avaliam as características de qualidade do produto (também conhecidas como variáveis de saída ou dependentes). A literatura tipicamente se refere às variáveis de processo como variáveis independentes, preditoras ou explanatórias, dentre outros termos; neste trabalho, variáveis de entrada serão referidas como independentes.

Os efeitos decorrentes de modificações de uma variável independente sobre outras variáveis (tanto independentes quanto dependentes) em estágios sucessivos podem ser de difícil avaliação. Kourti (2005) descreve fatores que promovem esta dificuldade: o caminho de processamento (que em alguns processos químicos pode não ser bem definido), tempos distintos de processamento em cada estágio e ocorrência de pontos de paradas entre estágios. Outro fator que dificulta a identificação do impacto das variáveis independentes sobre as características de qualidade desejadas do produto é a presença de variáveis correlacionadas (ZHANG *et al.*, 2010). A presença de correlação é fator restritivo à utilização de diversas técnicas multivariadas, dentre as quais a Regressão Linear Múltipla (RLM), visto que elevados níveis de correlação dificultam a estimativa precisa de coeficientes de regressão do modelo.

A determinação do relacionamento entre variáveis independentes e dependentes permite a identificação das condições ideais de operação de um processo multivariado organizado em diversos estágios (multiestágio). O que se busca então é a determinação de níveis ótimos das variáveis independentes que assegurem a obtenção dos valores desejados para as variáveis de resposta (ou valores próximos a esses) (JAECKLE e MACGREGOR, 1998; YACOUB e MACGREGOR, 2004).

Uma alternativa para determinação da condição ótima de trabalho consiste na aplicação de técnicas de variáveis latentes, que possuem a capacidade de análise em situações descritas por variáveis altamente correlacionadas, ruidosas, *outliers* (valores espúrios) e dados faltantes (WOLD *et al.*, 2001). A regressão por Mínimos Quadrados Parciais (PLS ou *Partial Least Squares*) se apoia em variáveis latentes e tem sido amplamente utilizada com vistas ao controle de processos multivariados. Tal regressão é detalhada em Jaeckle e MacGregor (1998) e Yacoub (2004) para processos contínuos de estágio único, e em Westerhuis e Coenegracht (1997), Jaeckle e MacGregor (2000), Chen e Tsai (2009) e Liu *et al.* (2011a) para processos em múltiplos estágios.

Observa-se que, em tais estudos, todas as variáveis independentes disponíveis são inseridas no modelo preditivo. Identifica-se aqui uma lacuna na literatura de controle multivariado de processos, pois a utilização de um grande número de variáveis independentes tipicamente incrementa a dificuldade na gestão do processo, principalmente aqueles em múltiplos estágios, podendo comprometer a precisão do modelo de previsão. De tal forma, a contribuição desta tese está na proposição de uma sistemática de seleção das variáveis independentes mais relevantes para construção de modelos robustos de regressão PLS e otimização dos níveis das variáveis independentes em ambientes multiestágio e com múltiplas variáveis de resposta. A grande maioria dos estudos no tema restringe-se a cenários descritos por uma única variável de resposta, normalmente associada a um único estágio produtivo.

## 1.1 TEMA

O tema desta tese é a seleção de variáveis independentes e determinação da condição ótima de operação em processos contínuos em múltiplos estágios e monitorados por múltiplas variáveis de resposta.

## 1.2 OBJETIVOS

O objetivo desta tese é desenvolver uma metodologia para seleção de variáveis independentes e identificação de condição ótima de operação para um processo contínuo multivariado e em múltiplos estágios.

Os objetivos específicos incluem:

- a) Propor um método de seleção de variáveis em contexto de regressão PLS aplicável a processos multivariados e em múltiplos estágios;
- b) Identificar condições ótimas de operação através de programação quadrática utilizando modelos de regressão PLS contendo variáveis independentes selecionadas;
- c) Analisar a aplicabilidade da sistemática em um sistema de produção de múltiplos estágios no setor avícola; e
- d) Analisar a robustez da sistemática proposta em um estudo de simulação.

### 1.3 JUSTIFICATIVA

Processos produtivos multivariados e em múltiplos estágios são comuns nas indústrias, sendo altamente complexos por conta do grande número de variáveis independentes utilizadas no seu controle. As técnicas tradicionais de controle de qualidade, que começaram a ser implantadas no Brasil na década de 80, não surtem efeito significativo no controle de tais processos, visto que as mesmas têm enfoque univariado. Desta forma, o presente estudo encontra respaldo prático na necessidade de desenvolvimento de abordagens aplicáveis contextos multivariados da indústria. De tal forma, objetiva-se o incremento na qualidade dos produtos através da obtenção das características nominais desejadas para os mesmos, melhor gerenciamento do processo, redução na variabilidade das variáveis de respostas, e maior conhecimento do processo através da identificação das variáveis independentes mais relevantes para caracterização de cada estágio do processo.

Pelo aspecto acadêmico, o presente trabalho justifica-se por apresentar uma nova sistemática de seleção de variáveis em processos multivariados e multiestágio, o que difere da maioria das abordagens sobre este tema ofertados pela literatura (as quais se focam em cenários compostos por uma única variável de resposta e geralmente limitados a processos com um único estágio de produção) (FRANK, 1987; FORINA *et al.*, 1999; HÖSKULDSSON, 2001; GAUCHI e CHAGNON, 2001; LAZRAQ *et al.*, 2003; CHONG e JUN, 2005 e MEHMOOD *et al.*, 2012). Salienta-se ainda a simplicidade do método proposto frente a outras abordagens da literatura, as quais tipicamente apoiam-se em ferramental complexo e pouco aplicável no contexto industrial.

A principal contribuição desta tese está na proposição de uma técnica de seleção de variáveis independentes em processos produtivos de múltiplos estágios. De posse de modelos reduzidos e informativos, é possível determinar as condições ótimas das variáveis independentes com vistas à minimização de desvios das características desejadas de qualidade (descritas por variáveis dependentes).

## 1.4 DELINEAMENTO DO ESTUDO

Considerando o objetivo proposto, define-se o método de pesquisa e método de trabalho.

### 1.4.1 Método de Pesquisa

A pesquisa, quanto à natureza, pode ser caracterizada como aplicada, uma vez que se propõe a gerar conhecimento para aplicações práticas, ou seja, soluções para problemas específicos (GIL, 2010). A abordagem é quantitativa, visto que pretende utilizar recursos e métodos estatísticos e linguagem matemática para descrever as relações entre as variáveis (GERHARDT E SILVEIRA, 2009). Em relação aos objetivos, a pesquisa se caracteriza como exploratória, pois envolve levantamento bibliográfico e análise de casos, o que permite a compreensão dos temas trabalhados (GIL, 2010).

Ainda segundo Gil (2010), pode-se caracterizar o delineamento da pesquisa desta tese como pesquisa bibliográfica e pesquisa experimental, uma vez que pretende identificar a influência das variáveis estudadas.

### 1.4.2 Método de Trabalho

O método de trabalho para a realização da pesquisa está dividido em quatro etapas, detalhadas na sequência.

A primeira etapa consiste no levantamento bibliográfico acerca de controle estatístico de processos multivariados, incluindo técnicas de seleção de variáveis para controle de processo multivariado, visando analisar as técnicas e resultados obtidos.

A segunda etapa baseia-se no estudo de métodos de projeção, principalmente o de mínimos quadrados parciais (PLS) e de ferramentas de otimização, como a programação quadrática.

A terceira etapa consiste na determinação da metodologia aplicada ao problema, compreendendo um critério adequado de seleção de variáveis, modelagem e posterior otimização do sistema.

A última etapa contém as análises e conclusões obtidas com a metodologia proposta.

## 1.5 DELIMITAÇÕES

O presente estudo está delimitado pelos métodos de modelagem orientada por dados (onde o PLS está incluído), não pretendendo abordar técnicas como métodos analíticos (espaços de estado ou fluxo de variação) ou métodos de engenharia, onde conhecimentos científicos do processo são considerados na gestão do processo.

Apesar de técnicas existentes em PLS abordarem relacionamentos não lineares, por exemplo, INLR (*Implicit non-linear latent variable regression*), não se pretende abordar estas técnicas.

Algumas suposições serão consideradas:

- Os erros de medidas são considerados desprezíveis;
- As amostras são obtidas, em cada estágio, no tempo correto, não havendo atraso ou avanço na coleta, o que poderia afetar as correlações entre as variáveis; e
- A identificação dos parâmetros das variáveis independentes não leva em consideração os custos decorrentes de perda de qualidade.

## 2 REFERENCIAL TEÓRICO

Este capítulo está dividido em oito seções, que tratam de tópicos necessários à compreensão do problema e à sua resolução. Primeiramente o setor avícola é abordado, com uma visão da sua importância, características e mercados, bem como com uma descrição da cadeia produtiva de aves; na sequência, é apresentada uma introdução sobre processos de múltiplos estágios; na terceira parte, é apresentada a utilização de métodos de projeção em controle de qualidade, e em seguida, é descrito o embasamento teórico da regressão PLS. Na seção seguinte são discutidos métodos de seleção de variáveis em PLS com propósitos de predição, e então conceitos de programação quadrática são apresentados; na sétima seção, uma discussão sucinta sobre índices de capacidade de processo e limites de especificação é proposta. Por fim, uma análise da função perda de qualidade de Taguchi é apresentada.

### 2.1 SETOR AVÍCOLA

O processo de múltiplos estágios analisado neste trabalho é de uma empresa do setor avícola do oeste de Santa Catarina. Este setor exerce impacto substancial na economia, não só na região oeste, mas na brasileira. A carne de frango foi o quinto item na pauta de produtos brasileiros exportados em 2013, com uma receita de quase 8 bilhões de dólares e um volume de 3,891 milhões de toneladas, de acordo com a União Brasileira de Avicultura (UBABEF, 2014), tendo sido exportada para mais de 150 países.

O Brasil mantém, desde 2004, a liderança na exportação mundial de carne de frangos, tendo perdido participação no mercado mundial nos últimos quatro anos, de 37% para 33,5% (AVISITE, 2014). O volume exportado de carne de peru em 2013 foi de 161 mil toneladas, gerando uma receita de 459 milhões de dólares e garantindo o segundo lugar no ranking mundial de exportação, ficando atrás somente dos Estados Unidos (Rural Br, 2014; UBABEF, 2013).

Os maiores produtores de carne de frango são Estados Unidos, China e Brasil, enquanto que, para perus, a ordem é Estados Unidos, Brasil e Canadá. A produção mundial de carne de frango em 2013 atingiu 82,178 milhões de toneladas, com 15,0% de participação da produção brasileira, enquanto que a produção de carne de peru em 2013 foi de 5,261 milhões

de toneladas, das quais 364 mil toneladas produzidas no Brasil, o que representa quase 7% da produção mundial (UBABEF, 2014).

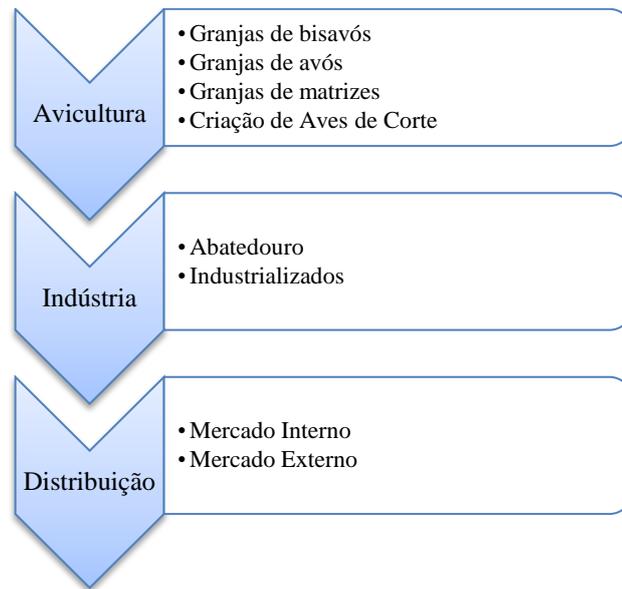
Em termos de produção, o número de frangos abatidos em 2013 totalizou 5,598 bilhões de cabeças, com um volume abatido de 12,346 milhões de toneladas, representando um crescimento de 7% em relação ao ano anterior (IBGE, 2014). A região Sul contribui com mais de 60% do número de frangos abatidos, tendo o estado do Paraná a maior participação, com 27,7%, seguido do Rio Grande do Sul, com 16,7% (ultrapassando Santa Catarina em 2013, cuja participação foi de 15,8%). Esta ordem é alterada ao considerar-se volume abatido, onde o Paraná permanece em primeiro, com 27,0%, Santa Catarina em segundo, com 17,3%, seguido do Rio Grande do Sul, com 15,5%. Na produção de perus, a região Sul participou com 65% da produção nacional; novamente o estado do Paraná teve a maior participação, com 26,6%, seguido de Santa Catarina, com 24,2% e do Rio Grande do Sul, com 14,3% (UBABEF, 2014).

Da produção nacional de frangos, 95% do volume é comercializado na forma in natura e somente 5% como industrializados. Do volume exportado, que corresponde a 31,6% da produção, os cortes de frangos (cortes de peito, coxa, sobrecoxa e asa, etc) são a maioria, com 53,1%, seguido dos frangos inteiros, com 38,1%, das carnes salgadas, com 4,6% e por último, dos industrializados, com 4,1%. Já em relação à carne de perus, 79% da produção é in natura e 21% industrializado. A exportação representa 44% do volume e os cortes de perus têm a maior participação, com 56,9%, seguido dos industrializados de perus, com 42,8% (UBABEF, 2014).

O setor avícola brasileiro emprega mais de 3,6 milhões de pessoas, de forma direta ou indireta, sendo que parcela deste contingente é de produtores integrados às empresas do setor, que fazem dos seus aviários a principal fonte de renda da propriedade (UBABEF, 2014). Esta condição é mais percebida na região Sul, onde os produtores integrados possuem pequenas propriedades, muitas vezes em locais de relevo acidentado, o que dificulta a utilização da área em outras atividades agrícolas ou pecuárias.

A cadeia produtiva de aves pode ser apresentada de diversas maneiras (LAZZARI, 2004; SANTINI, 2006; BASSI *et al.*, 2013), e neste trabalho, o foco está na cadeia principal (onde as aves estão envolvidas), não incluindo cadeias auxiliares (genética, nutrição, grãos, fabricação de premix e rações, medicamentos, assistência técnica, máquinas e equipamentos, aditivos, temperos e embalagens). A cadeia produtiva principal de aves está representada na Figura 1.

Figura 1 – Fluxograma dos segmentos produtivos da cadeia principal de aves



Fonte: elaborada pelo autor

O segmento de avicultura é o responsável pela multiplicação das aves e sua criação, abastecendo o segmento industrial. Esta multiplicação ocorre em várias etapas em ‘granjas’, as quais são formadas por um conjunto de aviários (também chamados de galpões).

As granjas de bisavós recebem pintos de um dia provenientes de granjas de *pedigrees* (a grande maioria destas granjas estão no exterior), onde são criadas as linhagens puras. No Brasil há granjas de bisavós de frangos, mas não de perus. Nas granjas de bisavós são alojadas normalmente quatro linhagens, combinadas com o objetivo de “produzir um casal de aves capaz de reproduzir em sua descendência o conjunto de características especiais dos seus ascendentes” (Jesus Jr. *et al.*, 2007). Uma vez alojadas nas granjas, as aves iniciam o seu crescimento, o qual tipicamente toma 24 semanas para frangos e requer um conjunto de aviários para sua operacionalização. Após esta idade, as aves estão maduras sexualmente, sendo transferidas para outros aviários, iniciando assim a fase de produção, onde as fêmeas colocam ovos em ninhos. Quando as aves chegarem à idade de 68 semanas para frangos, esta fase se encerra e as aves são abatidas (AVIAGEN, 2011).

As avós são a segunda geração, provenientes dos ovos de bisavós incubados e da mesma forma que as bisavós utilizam normalmente 4 linhagens. Os ovos são incubados em incubatórios e o tempo para completar o processo de incubação é de 21 dias para frangos e 28 para perus. Os tempos na fase de crescimento são de 24 semanas para frangos e 29 semanas

para perus. Na fase de produção, embora dependa da linhagem, a quantidade de ovos pode chegar a 160 por avó em postura, com uma eclosão média entre 75 e 79%; desta maneira, pode-se obter até 50 matrizes de um dia por avó. Nos perus a postura é menor, chegando a 110 ovos por avó, com eclosão média entre 81 e 84%, obtendo-se até 45 matrizes de um dia por avó em produção. Quando as aves chegarem à idade de 68 semanas para frangos e 60 para perus, esta fase se encerra e as aves seguem para o abate (AVIAGEN, 2011; HYBRID, 2011).

As granjas de matrizes recebem a terceira geração, composta somente por uma linhagem. Os tempos de crescimento e de produção são os mesmos das avós, mas o desempenho pode alcançar, em frangos, 190 ovos por matriz em postura e, após incubação, 150 pintos por matriz. Para perus, geram-se 110 ovos por matriz e, após incubação, 90 perus por matriz (AVIAGEN, 2011b; HYBRID, 2011).

A última etapa no segmento avicultura é a criação de aves de corte, que geralmente ocorre em produtores rurais, parceiros integrados da empresa avícola. Neste sistema de parceria, os integrados se comprometem a fornecer aviários adequados (com cama, aquecimento, iluminação e água) e mão de obra, enquanto que a empresa fornece pintos de um dia, rações, medicamentos, transportes e assistência técnica, além de remunerar o integrado. O tempo desta fase para frangos é de 28 a 50 dias, conforme o peso desejado no abate, enquanto que para perus normalmente é de 56 a 174 dias e pode ser dividido em duas fases, iniciador (cerca de 25 dias) e terminador (até o abate).

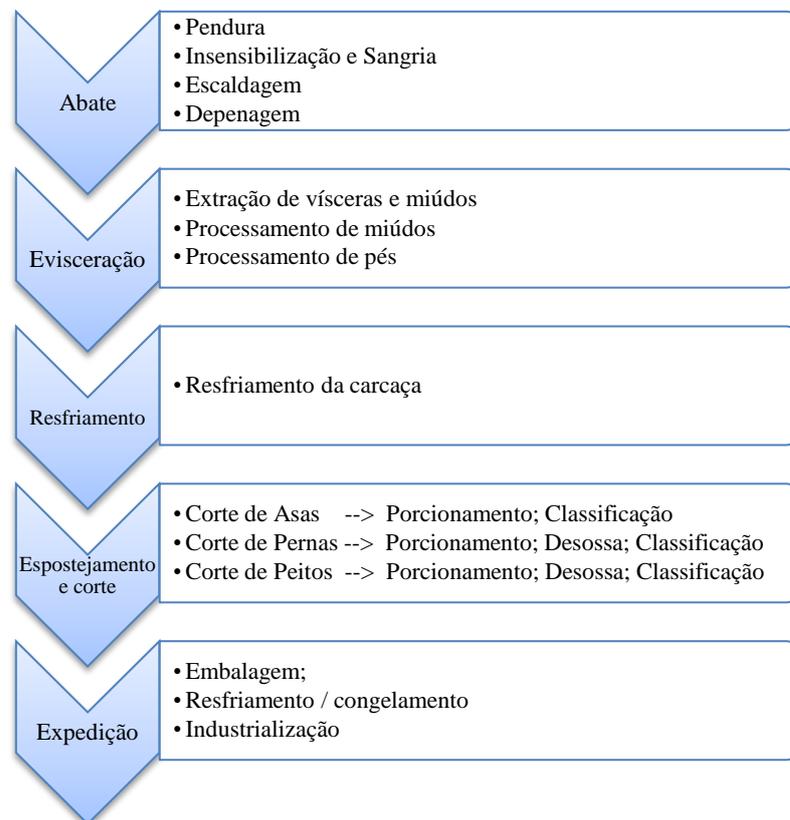
Observam-se assim duas características importantes neste segmento sob a ótica de produção: tempo de ciclo e multiplicação. O tempo do ciclo é longo, pois, a partir do alojamento de bisavós de um dia (de frangos), os primeiros frangos a serem abatidos oriundos destas bisavós chegarão ao abatedouro em 88 semanas e os últimos em 223 semanas, ou seja, mais de 4 anos após. Nos perus, considerando somente o alojamento de avós de um dia, os primeiros perus levarão 82 semanas para chegar ao abatedouro e os últimos 155 semanas, ou 3 anos. No efeito multiplicação de uma bisavó de um dia (de frangos), pode-se chegar a 280 mil frangos abatidos, enquanto que, para uma avó de peru de um dia, pode-se esperar 2,5 mil perus abatidos.

A cadeia principal, apresentada na Figura 1, pode estar incompleta em pequenas e médias empresas, pois algumas empresas não possuem granjas de bisavós, de avós, ou até mesmo granjas de matrizes, optando por comprar pintos com idade de um dia.

O segundo segmento da cadeia abrange a área industrial, inicialmente pelo abatedouro (ou frigorífico), seguido da industrialização. No abatedouro geralmente são beneficiados frangos (ou perus) inteiros ou cortes (partes), considerados produtos *in natura*, enquanto que nos industrializados há diversas famílias de produtos, tais como empanados, embutidos, cozidos, marinados, defumados, temperados e grelhados.

A Figura 2 apresenta as etapas do processo num abatedouro típico para produção de cortes. Estas etapas estão descritas (parcialmente) nos trabalhos de Vegro e Rocha (2007), Espíndola (2009), e Voilà e Triches (2013). Os processos podem variar ligeiramente conforme o animal e o nível de automação empregada, bem como com o tipo de produtos fabricados. Tais etapas são detalhadas na sequência.

Figura 2 - Fluxograma do processamento de aves



Fonte: Elaborada pelo autor

Iniciando pelo abate, segue uma descrição sucinta de cada etapa:

a) **Pendura:** As aves chegam transportadas por caminhões dos aviários de criação até o abatedouro, em gaiolas ou em caixas de transporte; as aves são então retiradas e penduradas pelas patas num transportador aéreo denominado nória.

b) **Insensibilização:** Após a pendura, as aves seguem pela nória para a insensibilização, onde podem receber um banho de água na cabeça e um choque elétrico ou passar por uma câmara de atmosfera controlada, geralmente com dióxido de carbono. O objetivo é proporcionar um estado de insensibilidade, porém mantendo as funções vitais até a sangria.

c) **Sangria:** Imediatamente após a insensibilização acontece o abate, com um corte profundo no pescoço das aves (jugulares); ocorre então o escoamento do sangue, antes que o animal recupere a sensibilidade.

d) **Escaldagem:** É realizada num tanque de água quente (cerca de 60°), geralmente em contrafluxo, com o objetivo de lavagem, afrouxamento da fixação e aglutinação das penas. A ocorrência de temperaturas mais altas ou maiores tempos de permanência em contato com a água pode provocar queimaduras e rompimentos na pele, afetando o aproveitamento da peça.

e) **Depenagem:** Ocorre por meio de dedos de borrachas, que devem agir sobre toda a superfície da ave, retirando todas as penas. A distância dos dedos até a ave deve ser adequadamente ajustada; caso a distância seja curta, pode-se gerar rompimento de pele, presença de hematomas e ossos quebrados; se a distância for longa, pode-se verificar a presença de penas na carcaça.

f) **Extração de vísceras e miúdos:** Esta etapa pode ser automática ou manual, sendo responsável pelo corte dos pés, extração da cloaca, abertura do abdômen, extração do pacote de vísceras, papo e traqueia, separação de miúdos e corte de pescoço. Os miúdos separados (coração, fígado e moela), assim como os pés e pescoço, são direcionados para processamento (o qual inclui resfriamento, toalete, classificação, embalagem e congelamento).

g) **Resfriamento:** Após a evisceração, as carcaças devem ser resfriadas para garantia da qualidade do produto (durabilidade e contaminação). O resfriamento pode acontecer por imersão em água gelada ou em túneis de ar. Concluído o resfriamento, a carcaça pode seguir um caminho diferente, dependendo do tipo de produto a ser fabricado. Se o produto é frango inteiro, a carcaça segue para embalagem e congelamento (caso não seja frango resfriado); se o produto se destina a cortes, a carcaça segue para espostejamento.

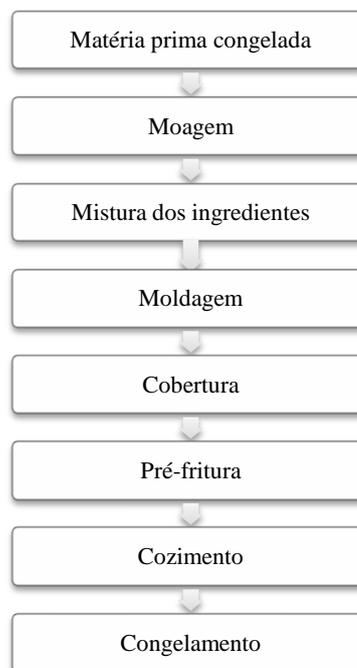
h) **Espostejamento:** A carcaça é cortada e suas partes separadas em asas, pernas e peito. As asas podem então ser separadas em até três partes: ponta, meio e coxinha da asa, enquanto que as pernas podem ser separadas em coxa e sobrecoxa (as quais também podem ser desossadas); e, por último, o peito pode ser obtido inteiro (com ou sem osso) ou em meio-peito (com ou sem pele). As peças podem ser classificadas, embaladas e congeladas, ou

enviadas para industrialização. Após o esposteamento, a carcaça está reduzida ao dorso e a carne ainda existente é separada mecanicamente.

i) **Expedição:** Os produtos congelados e paletizados seguem para a expedição, onde são armazenados em câmaras de congelamento até o embarque.

Na industrialização, cada família de produtos utiliza um processo distinto, composto por diversos equipamentos (tais como formas, gaiolas, tanques, injetoras, embutideiras, pasteurizadores, estufas, fritadeiras, porcionadoras); tais equipamentos não são detalhados neste estudo. Como exemplo, a Figura 3 apresenta o fluxograma de processamento de produtos empanados; mais informações sobre as etapas de industrialização podem ser encontradas em Sarcinelli *et al.* (2007) e Dill *et al.* (2009).

Figura 3 - Fluxograma do processo de produtos empanados



Fonte: Adaptada de Dill *et al.* (2009)

A partir da breve descrição do processo feita acima, percebe-se que a cadeia produtiva do setor avícola é extensa e diversa, sendo que em cada etapa há um grande número de variáveis (independentes e de resposta). Soma-se a isso a existência de correlações de difícil avaliação, as quais incluem aspectos ligados a questões sanitárias, de higiene e grau de incerteza no desempenho zootécnico. Tais condições mostram-se propensas à aplicação da sistemática proposta neste estudo.

## 2.2 PROCESSOS EM MÚLTIPLOS ESTÁGIOS

Um processo multiestágio é um sistema de múltiplos componentes, estações ou estágios necessários para finalizar um produto ou serviço, tendo como características principais: (i) existência de estágios múltiplos e estruturas com configurações sequenciais, paralelas ou mistas; e (ii) suposição de que resultados de um estágio constituem as entradas do próximo estágio e que a variação ocorrida neste estágio é adicionada à variação dos estágios anteriores. Estabelece-se então um sistema coletivo, onde o desempenho final depende da variação acumulada em cada estágio (SHI e ZHOU, 2009).

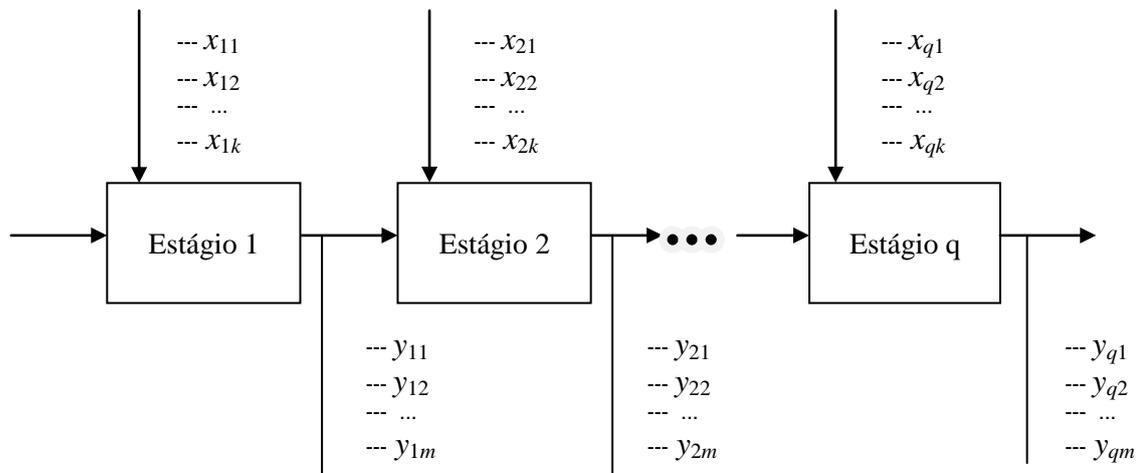
Esse tipo de processo é vastamente encontrado em ambientes industriais, principalmente em processos químicos, farmacêuticos e mecânicos, podendo ser do tipo contínuo ou em bateladas. Células de manufatura e sistemas de cadeia de abastecimento também podem ser enquadradas neste tipo de processo (ÜNDEY e ÇINAR, 2002; SHI e ZHOU, 2009).

A Figura 4 ilustra um processo em cascata multiestágio onde a entrada se dá por meio das variáveis independentes ( $x'_s$ ) e a saída pelas variáveis de resposta ( $y'_s$ ). Uma variável independente pode influenciar o valor de uma variável de resposta no próprio estágio em que se encontra, bem como nos estágios posteriores em função do efeito cascata.

Neste estudo, cada variável ( $x$  ou  $y$ ) utiliza um subíndice  $q$  para definir o estágio onde esta variável se localiza, e um segundo subíndice  $j$  para indicar o número desta variável neste estágio. Cada estágio é descrito por  $k$  variáveis independentes e  $m$  variáveis de resposta (podendo  $k$  e  $m$  assumirem valores distintos por estágio). Seguindo esta definição, uma variável  $x_{13}$  é a terceira variável do primeiro estágio. Na Figura 4, a variável  $x_{21}$  (pertencente ao segundo estágio) pode impactar em qualquer variável de qualidade deste estágio ( $y_{2j}$ ) até o último estágio ( $y_{qj}$ ), mas não pode gerar nenhuma alteração nas variáveis de resposta do primeiro estágio ( $y_{1j}$ ), de acordo com o efeito cascata (HAWKINS, 1993).

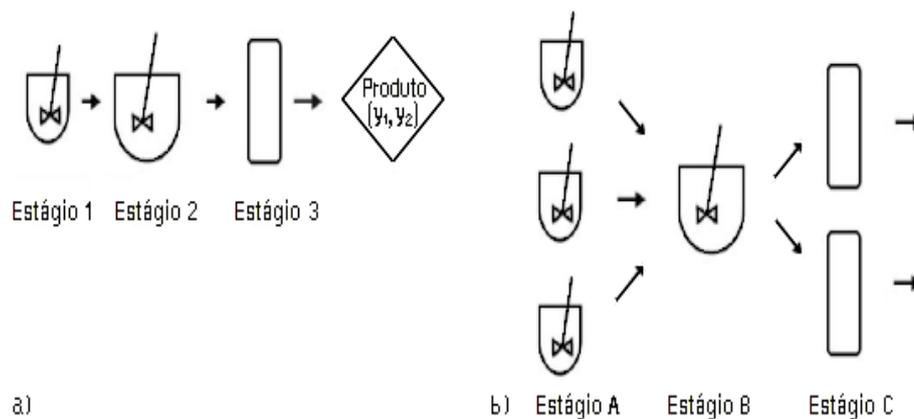
Em função do tipo de configuração do processo multiestágio, pode ser difícil identificar claramente o fluxo de processamento (KOURTI, 2003a). Na Figura 5a, três estágios estão colocados em série e o produto passa por cada um deles de acordo com uma sequência estabelecida; na Figura 5b, por sua vez, o primeiro estágio é composto por várias operações idênticas em paralelo, sendo agrupadas em uma única operação no segundo estágio e então distribuída em duas operações em paralelo no estágio seguinte.

Figura 4- Exemplo de um processo em cascata multiestágio



Fonte: Adaptada de JEARKPAPORN *et al.* (2007)

Figura 5 - Exemplo de fluxos de processamento em processos multiestágios



Fonte: Adaptado de KOURTI (2003a)

Um critério importante em processos em múltiplos estágios consiste na determinação dos limites de um estágio (seu início e término). Conhecimento científico e análise de dados são abordagens que possibilitam tal fim. A primeira considera a presença de operações unitárias distintas em cada estágio e assume-se, por conhecimento de especialistas, que após determinada operação há variação significativa em uma ou mais características do produto. Por sua vez, a segunda apoia-se em análise do processo, identificando pontos onde há mudanças na estrutura de correlação do processo (CAMACHO *et al.*, 2008; YAO e GAO, 2009).

### 2.3 UTILIZAÇÃO DE MÉTODOS DE PROJEÇÃO EM CONTROLE DE QUALIDADE

O desenvolvimento de abordagens apoiadas em variáveis latentes tem contribuído significativamente no controle de processos multivariados, pois possibilitam a construção de modelos de dimensão reduzida para processos com grande número de variáveis.

Kourti (2005) descreve os principais avanços ocorridos no monitoramento de processos e em técnicas de detecção de falhas, enquanto que uma visão mais abrangente é dada por MacGregor e Cinar (2012), que classificam os modelos conforme o objetivo. Tais objetivos são sucintamente apresentados abaixo, apontando estudos significativos em cada frente.

a) **Monitoramento do processo:** As variáveis de resposta (parâmetros do produto) e as variáveis independentes (parâmetros do processo) são coletadas com a suposição de estarem ‘sob controle’; é então criado um modelo e as observações futuras são comparadas a este modelo para verificar sua aderência (BERSIMIS *et al.*, 2007). O conceito do monitoramento está baseado em cartas de controle, geralmente utilizando uma estatística como o  $T^2$  de Hotelling, que avalia a distância entre o valor observado e o previsto pelo modelo. O uso de métodos de projeção ganha importância quando o número de variáveis a ser monitoradas aumenta, uma vez que tais métodos reduzem a dimensão do problema, facilitando sua interpretação.

Como trabalhos relevantes para o monitoramento de processo, pode-se citar Jackson (1991), que utilizou a Análise de Componentes Principais (PCA), e Nomikos e MacGregor (1994), que criaram o método PCA Multidirecional (MPCA), no qual observações são ordenadas no tempo, formando um bloco tridimensional. Por sua vez, Skagerberg *et al.* (1992) mostraram que a regressão PLS (detalhada na seção 2.4) pode ser usada no monitoramento de processo por meio de uma simulação de processo; os autores observaram que o método permite estabelecer bons modelos preditivos, os quais podem configurar esquemas de controle inferencial. Com propósitos semelhantes, MacGregor *et al.* (1994) trataram cada seção do processo como um bloco e assim analisaram as variáveis latentes derivadas de cada bloco; por fim, analisaram o conjunto de blocos que caracteriza o processo como um todo. Nomikos e MacGregor (1995) propuseram ainda a metodologia PLS Multidirecional (MPLS) aplicada ao monitoramento de processos em bateladas.

b) **Diagnóstico do processo:** Num segundo momento, após o monitoramento sinalizar uma situação ‘fora de controle’ (ou seja, uma falha no processo), deve-se identificar as variáveis independentes responsáveis pela situação (RAICH e ÇINAR, 1996). Para tanto, pode-se avaliar a variação no erro quadrático de predição calculado, antes e imediatamente após a sinalização de ‘fora de controle’, com os valores preditos determinados pelo modelo PLS. Pode-se então calcular a contribuição de cada variável no erro de predição (KOURTI e MACGREGOR, 1995). Maiores detalhes sobre abordagens voltadas ao diagnóstico de processo podem ser encontrados em Venkatasubramanian *et al.* (2003), MacGregor *et al.* (2005), Kourti (2005) e Qin (2012).

c) **Otimização e controle do processo:** O terceiro objetivo é o de otimização, onde se busca encontrar as condições ótimas de operação do processo, minimizando os desvios dos valores desejados nas variáveis de resposta. MacGregor *et al.* (1994), utilizando um modelo MBPLS (PLS em múltiplos blocos), propuseram uma metodologia que possibilita gerar um produto com características desejadas, de tal forma que as variáveis independentes permaneçam dentro de uma ‘janela de condições recomendadas de operação’. Inicialmente, os autores analisaram um conjunto de dados históricos composto por medidas das características de qualidade (variáveis de resposta) juntamente com as condições das variáveis de processo (independentes). Na sequência, construíram um modelo para caracterizar o comportamento dos dados do processo e o relacionamento existente entre as condições do processo e as características de qualidade; desta forma, encontraram novas condições do processo para um produto com características de qualidade desejadas.

Outra abordagem foi proposta por Jaeckle e MacGregor (1998), a qual apoia-se na inversão do modelo (estimando  $\hat{x}$  ao invés de  $\hat{y}$ ). Nesta metodologia, compara-se o número de variáveis de qualidade desejadas  $k$  com o número de variáveis latentes  $A$ ; no caso mais comum, quando  $k$  é menor do que  $A$ , infere-se a existência de um conjunto infinito de soluções (ou seja, várias condições do processo que permitirão chegar às características desejadas de qualidade). Os autores salientam a importância dos parâmetros de processo estimados estarem dentro das condições observadas no processo, de forma que os valores previstos para as variáveis independentes tenham sido observadas no histórico. Também é importante garantir que estas condições de processo mantenham a mesma estrutura de covariância dos dados históricos. Segundo Burnham *et al.* (1999), estas condições precisam ser válidas também para novos dados. Em trabalho posterior, Jaeckle e MacGregor (2000) aplicaram a metodologia proposta em um processo de polimerização em bateladas de dois

estágios: o primeiro composto pelos ingredientes utilizados no início do processo e, após um período de reação, a adição de novos ingredientes caracteriza o segundo estágio.

Para processos em bateladas, Kourti (2003b) avançou nos conceitos de transição e alinhamento de trajetórias de variáveis. O alinhamento é uma maneira de tratar bateladas que apresentaram durações diferentes entre si, enquanto que a transição está relacionada à mudança na condição do processamento (tais como início e reinício do processo). Flores-Cerrillo e MacGregor (2004) recalcularam as trajetórias de variáveis em diversos pontos durante uma batelada; para tanto, utilizaram um modelo PLS e realizaram a inversão do modelo de acordo com as proposições de Jaeckle e MacGregor (1998) para prever as variáveis de resposta. García-Muñoz *et al.* (2003) avançaram neste tema em processos multiestágios, criando variáveis para alinhamento de trajetórias em cada estágio.

Yacoub e MacGregor (2004) apresentaram uma metodologia para monitoramento e controle de processos baseada na otimização do subespaço criado por um modelo de variáveis latentes, utilizando a abordagem de inversão de modelos introduzida por Jaeckle e MacGregor (1998) e incorporando restrições às variáveis de processo. Para tanto, integraram uma regressão PLS não linear a um modelo de Programação Quadrática Sequencial. A abordagem proposta apoiou-se nas seguintes premissas (i) que as condições geradas fossem consistentes com as condições de operação passadas; (ii) que as variáveis de resposta fossem alcançadas, e (iii) que as condições encontradas resultassem em pequenas variâncias nas variáveis de resposta. Os autores enfatizam que suas proposições podem ser utilizadas em duas frentes distintas: novo produto, onde objetiva-se encontrar condições de processo capazes de produzir um novo produto com a menor variância possível, e controle da qualidade do produto, visando tratar de situações de ‘fora de controle’ e enumerar ações para recuperar o controle do processo.

Com propósitos semelhantes, Muteki *et al.* (2006) apresentaram uma abordagem para desenvolvimento de novos produtos, observando a seleção de matérias-primas e a sua proporção. Utilizando PLS seguida pela inversão do modelo de Jaeckle e MacGregor (1998), a otimização foi operacionalizada com vistas à minimização do custo e do número de matérias-primas envolvidas no processo, bem como dos desvios em relação às características desejadas do produto. Tomba *et al.* (2012) apresentaram uma estrutura geral para a inversão de modelos de variáveis latentes (MRVL) para o projeto de novos produtos e definição de condições de um processo produtivo. Para tanto, sugerem os seguintes passos: a) Construir um modelo MRVL entre as matrizes de variáveis independentes e de resposta; b) Determinar

as especificações desejadas do produto, isto é, especificações das variáveis de resposta, por meio de valores atribuídos (restrições de igualdade), restrições uni ou bilaterais (desigualdade) e limites físicos; c) Determinar as restrições necessárias para as variáveis independentes (se houver) e o tipo de restrição (igualdade, desigualdade e limites físicos), tal que a solução encontrada seja de relevância prática; d) Verificar se as especificações das variáveis de resposta são coerentes com os dados históricos; caso sejam coerentes, inverter o modelo, caso não o sejam, não inverter o modelo; e e) Otimizar o sistema, apresentando as previsões para as variáveis de resposta, estatística  $T^2$  de Hotelling e erros de predição.

Face às abordagens acima descritas, percebe-se que limitados estudos têm se focado na otimização e controle de processos com múltiplos estágios através de variáveis latentes. Um dos trabalhos pioneiros foi o de Westerhuiss e Coenegracht (1997), os quais visaram prever as variáveis de resposta e melhor compreender um processo produtivo de comprimidos farmacêuticos de dois estágios. Para tanto, utilizaram um modelo MBPLS com dois blocos, um endereçando as variáveis de processo (dos dois estágios) e a composição da mistura de pó, enquanto que o outro continha as propriedades físicas dos granulados.

Recentemente, Chen e Tsai (2009) propuseram duas estratégias de controle para processos de múltiplos estágios através de um modelo MBPLS. Na primeira estratégia, após cada estágio, é feita uma reotimização do modelo e nova predição para as variáveis dos estágios seguintes, enquanto que na segunda este procedimento é executado após o término do processo. Por fim, em Liu *et al.* (2011b) é apresentada uma abordagem para inversão do modelo de regressão de variáveis latentes em um processo multiestágio para a fabricação de comprimidos. Aplica-se inicialmente um modelo MBPLS com o objetivo de identificar as seções da linha de fabricação mais influentes para a qualidade final do produto. Os blocos são formados por frações de massa de API (ingredientes farmacêuticos ativos) e excipientes usados, variáveis independentes (de processo), configurações da prensa de comprimidos e variáveis de resposta. A otimização é realizada sobre o espaço das variáveis latentes do modelo PLS.

## 2.4 REGRESSÃO PARTIAL LEAST SQUARES (PLS)

A regressão Partial Least Squares (PLS) permite modelar o relacionamento entre grupos de variáveis de resposta e variáveis independentes, buscando a máxima covariância

entre as combinações lineares (variáveis latentes) destes dois grupos de variáveis (WOLD *et al.*, 2001; ABDI, 2010). Para Wold *et al.* (2001), variáveis latentes permitem prever as variáveis de resposta  $\mathbf{Y}$ , além de modelarem as variáveis independentes  $\mathbf{X}$ . Conforme Lattin *et al.* (2011), variáveis latentes não são diretamente observadas nos modelos de medida, mas podem ser descritas como funções lineares das variáveis observadas (FRANK e KOWALSKI, 1984).

Dentre suas vantagens, a regressão PLS permite analisar sistemas descritos por variáveis altamente correlacionadas, com um grande número de variáveis independentes e/ou variáveis de resposta, e com dados faltantes (WOLD *et al.*, 2001). Detalhes acerca da regressão PLS podem ser encontrados em Höskuldsson (1988), Wold *et al.* (2001), Rosipal e Krämer (2005) e Abdi (2010), bem como em Geladi e Kowalski (1986). A seguir estão descritos os fundamentos matemáticos da regressão PLS.

Seja um conjunto de dados dispostos em duas matrizes, uma de variáveis independentes, denominada matriz  $\mathbf{X}$ , e uma matriz  $\mathbf{Y}$  com as variáveis de resposta, sendo  $\mathbf{X}$  composta por  $k$  variáveis (colunas),  $\mathbf{Y}$  por  $m$  variáveis e ambas por  $n$  observações (linhas). Cada observação do processo (amostra) pode ser representada por um vetor  $\mathbf{x}_i$  ( $x_{i1}, x_{i2}, \dots, x_{ik}$ ); por sua vez, uma observação dos dados de qualidade do produto pode ser representada por um vetor  $\mathbf{y}_i$  ( $y_{i1}, y_{i2}, \dots, y_{im}$ ).

As variáveis latentes são formadas por combinações lineares entre as variáveis originais da matriz  $\mathbf{X}$ , associando um  $w$  a cada variável, conforme a equação (1). A representação matricial é colocada em negrito.

$$t_{ia} = w_{1a}x_{i1} + w_{2a}x_{i2} + \dots + w_{ka}x_{ik} = \mathbf{w}'\mathbf{x} \quad (1)$$

Os pesos  $w$ , armazenados no vetor  $\mathbf{w}$ , indicam a importância de cada variável  $x$ ; variáveis com peso de alto valor numérico implicam que são importantes (WOLD *et al.*, 2001). As variáveis latentes  $\mathbf{t}$  são chamadas de escores de  $\mathbf{X}$ . Uma característica importante é que os escores  $\mathbf{t}$  são poucos ( $t_a = 1, 2, \dots, A$ ) e ortogonais uns aos outros (WOLD *et al.*, 2001).

De forma análoga, variáveis latentes também são formadas por combinações lineares das variáveis que compõem a matriz  $\mathbf{Y}$ , às quais é associado um peso ( $\mathbf{c}$ ), conforme a equação (2). Por analogia, as variáveis latentes  $\mathbf{u}$  são denominadas escores de  $\mathbf{Y}$ .

$$u_{ia} = c_{1a}y_{i1} + c_{2a}y_{i2} + \dots + c_{ma}y_{im} = \mathbf{c}'\mathbf{y} \quad (2)$$

Os escores  $\mathbf{t}$  são utilizados como estimadores da matriz  $\mathbf{X}$ , pois com o produto dos escores  $\mathbf{t}$  pela cargas  $\mathbf{p}$  se obtém uma matriz que se aproxima de  $\mathbf{X}$  e cuja diferença residual reduz-se com o aumento dos componentes retidos. As cargas  $\mathbf{p}$  são os coeficientes de regressão entre as variáveis independentes sobre os escores  $\mathbf{t}$  (VANDEGINSTE, 1998). A predição de um elemento da matriz  $\mathbf{X}$  pode ser representada conforme a equação (3). O termo  $e$  na equação (3) representa o resíduo entre o valor observado e o estimado.

$$x_{ik} = t_{i1}p_{1k} + t_{i2}p_{2k} + \dots + t_{ia}p_{ak} + e_{ik} = \mathbf{t}'\mathbf{p} + \mathbf{e} \quad (3)$$

O mesmo procedimento pode ser operacionalizado na matriz  $\mathbf{Y}$ , utilizando-se dos escores de  $\mathbf{Y}$  ( $\mathbf{u}$ ) e das cargas  $\mathbf{c}$ , conforme a equação (4); o resíduo é representado pelo termo  $\mathbf{g}$ .

$$y_{im} = u_{i1}c_{1m} + u_{i2}c_{2m} + \dots + u_{ia}c_{am} + g_{im} = \mathbf{u}\mathbf{c} + \mathbf{g} \quad (4)$$

Na sequência, definem-se os coeficientes da regressão PLS,  $b_{mk}$  ( $\mathbf{b}$  no formato vetorial), conforme a equação (5), os quais permitem realizar predições das variáveis dependentes conforme a equação (6). No caso de existir apenas uma variável de resposta,  $m=1$ .

$$b_{mk} = \sum_a c_{ma} \sum_k w_{ka} = \mathbf{c}'\mathbf{w} \quad (5)$$

$$y_{im} = \sum_k b_{mk} x_{ik} + f_{im} = \mathbf{b}\mathbf{x} + \mathbf{f} \quad (6)$$

Antes de executar a regressão PLS, recomenda-se padronizar os dados, reduzindo efeitos de magnitude das variáveis e facilitando a comparação entre elas. Se necessário, pode-se ainda transformar a escala de uma variável quando o objetivo for tornar a sua distribuição mais simétrica. Se a amplitude da magnitude dos dados for superior a dez, recomenda-se uma transformação logarítmica (WOLD *et al.*, 2001).

O processo de obtenção dos componentes principais é iterativo: os vetores  $\mathbf{t}$ ,  $\mathbf{u}$ ,  $\mathbf{w}$ ,  $\mathbf{c}$  e  $\mathbf{p}$  são determinados para a primeira variável latente, a qual explica a maior parte da covariância entre  $\mathbf{X}$  e  $\mathbf{Y}$ ; as variáveis latentes subsequentes explicam sistematicamente menores quantidades da covariância. Isto ocorre devido ao processo de ‘deflacionamento’, no qual uma variável latente, depois de formada, tem o seu impacto subtraído da matriz  $\mathbf{X}$ . Assim, as últimas variáveis latentes tipicamente não contribuem de forma significativa para a previsão de  $\mathbf{Y}$ , podendo ser descartadas e reduzindo a dimensionalidade da matriz  $\mathbf{X}$  (GELADI e KOWALSKI, 1986; HÖSKULDSSON, 1988; ABDI, 2010). O algoritmo NIPALS (*Non-linear Iterative Partial Least Squares*), utilizado neste trabalho para

determinação dos parâmetros da regressão PLS, é detalhado em Geladi e Kowalski (1986); este algoritmo é dito ‘não linear’, pois maximiza a covariância entre as matrizes  $\mathbf{X}$  e  $\mathbf{Y}$  através dos escores  $\mathbf{t}$  e  $\mathbf{u}$ . Entretanto, o relacionamento entre as variáveis latentes é modelado por uma função linear e, portanto, adequado para tratar de modelos de regressão PLS linear (ROSIPAL e KRÄMER, 2005).

Assim, um ponto importante a ser definido em PLS é o número de componentes  $A$  a serem retidos (GELADI e KOWALSKI, 1986; GARTHWAITE, 1994; WOLD *et al.*, 2001). Dentre os vários critérios existentes para definir o número de componentes, destacam-se o limite mínimo da norma da matriz de resíduos e soma dos quadrados dos resíduos de predição (PRESS) em Geladi e Kowalski (1986), e validação cruzada em Wold *et al.* (2001) e Höskuldsson (1988).

A validação cruzada é o método mais utilizado na determinação dos erros de predição e tornou-se padrão na análise de regressão PLS (HASTIE *et al.*, 2009, WOLD *et al.*, 2001). De acordo com Lattin *et al.* (2011), a validação cruzada serve para testar o modelo quanto à sua capacidade preditiva, sendo realizada mediante a divisão do conjunto de dados, de forma aleatória, em duas porções: treino e teste. A porção de treino é utilizada para calibrar o modelo PLS, ao passo que as observações no conjunto de teste (representando novas observações) são preditas pelo modelo gerado. A validação é feita estimando-se os resíduos da porção de teste. A quantidade de observações deve ser maior na porção de treino do que na de teste: Lattin *et al.* (2011) propõem dois terços e um terço, respectivamente, enquanto que Wold *et al.* (2001) e Hastie *et al.* (2009) sugerem dividir o conjunto de dados por um valor inteiro de cinco a nove, o que representa uma proporção entre oitenta a noventa por cento das observações para a porção de treino.

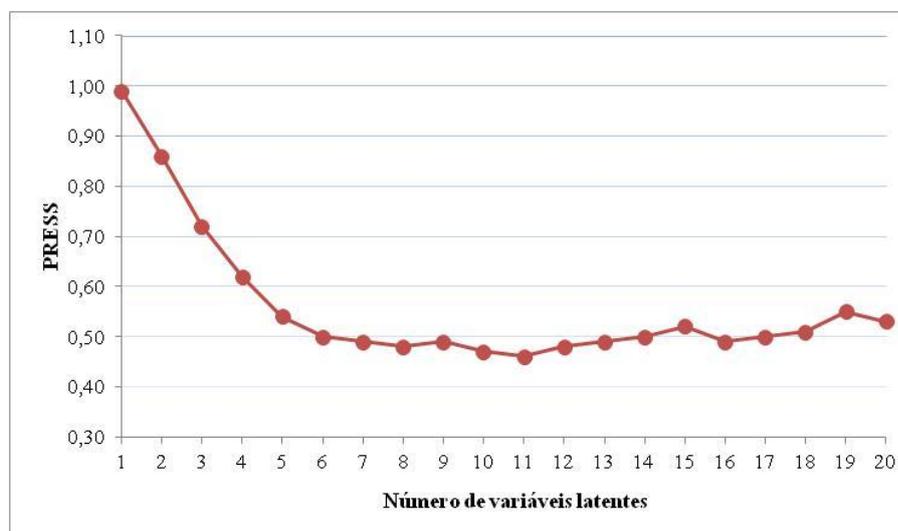
No sistema de validação cruzada total, descrito em Wold *et al.* (2001), avalia-se a capacidade de predição do modelo por meio da estatística PRESS (*predictive residual error sum of squares* ou soma dos quadrados dos resíduos de predição) que mede o erro de predição, conforme a equação (7). Desta maneira, se o conjunto de dados for dividido por cinco, por exemplo, o número de modelos criados também será cinco.

$$\text{PRESS}_i = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (7)$$

onde  $y_i$  é o valor da variável para a  $i$ -ésima observação;  $\hat{y}_i$  é o valor predito para  $y_i$  e  $n$  é o número de observações.

A escolha do número de componentes pode ainda ser realizada através de uma análise gráfica, confrontando-se o critério de avaliação do erro (PRESS ou outro critério) pelo número de variáveis latentes utilizados no modelo. A decisão usual consiste na retenção do número de componentes que minimiza o erro. Diversos autores, no entanto, sustentam que é melhor aceitar um pequeno incremento no erro ao trabalhar-se com um número maior de componentes, como ilustrado na Figura 6 (GELADI e KOWALSKI, 1986; VANDEGINSTE *et al.*, 1998; HASTIE *et al.*, 2009). No exemplo desta figura, o número de componentes retidos é seis, enquanto que o ponto mínimo na linha ocorre com onze variáveis latentes.

Figura 6 - Perfil hipotético na determinação do número de variáveis latentes



Fonte: Adaptado de GELADI E KOWALSKI (1986).

Kim *et al.* (2007) argumentam que, quando as variáveis de resposta são coletadas de atividades distintas e influenciadas por diferentes conjuntos de variáveis independentes, os modelos PLS tendem a ter muitos componentes, o que pode dificultar a interpretação. De maneira semelhante, se um grupo de variáveis de respostas é majoritariamente influenciado por um grupo restrito de variáveis independentes, a interpretação de causalidade fica comprometida.

## 2.5 SELEÇÃO DE VARIÁVEIS EM REGRESSÃO PLS

Algumas deficiências na regressão PLS, descritas por Mehmood *et al.* (2012), ocorrem quando há um grande número de variáveis e poucas observações (o que pode prejudicar a

modelagem por conta de problemas na consistência dos estimadores), e quando há muitas variáveis independentes irrelevantes, o que pode gerar grande variação na predição das variáveis de resposta.

Os objetivos da seleção de variáveis no contexto da regressão PLS incluem: melhorar o desempenho de previsão, identificar preditores mais eficientes em termos de custo e melhorar o entendimento do processo (GUYON e ELISSEEFF, 2003). Os diferentes métodos de seleção de variáveis reportados pela literatura evidenciam um aumento substancial na capacidade preditiva do modelo PLS quando um conjunto reduzido de variáveis independentes é utilizado (FORINA *et al.*, 1999; GAUCHI e CHAGNON, 2001; HOSKULDSSON, 2001; LAZRAQ *et al.*, 2003).

Andersen e Bro (2010) listam diversos aspectos a serem endereçados em procedimentos de seleção de variáveis em PLS com propósitos de predição. O primeiro é o sobreajuste (*overfitting*), o qual pode ocorrer quando a proporção de variáveis em relação ao número de observações for grande. O modelo encontra algumas variáveis que se correlacionam com a variável de resposta, porém com reduzida relação de causa e efeito, fazendo com que as previsões não apresentem consistência. Para contornar este problema, Andersen e Bro (2010) recomendam a utilização de um processo de validação cruzada na determinação do modelo.

O segundo é a redundância, na qual considera-se que, se duas ou mais variáveis independentes têm um comportamento muito semelhante e estão correlacionadas com uma variável de resposta, o modelo PLS terá dificuldade em identificar qual delas é a responsável pela variação da variável de resposta (supondo que há uma relação de causa e efeito). Tendo em vista que tais variáveis independentes são redundantes ao modelo, a eliminação de uma delas não melhora a qualidade da previsão, porém torna o modelo menor e mais simples de interpretar; contudo, a remoção de variáveis redundantes pode ser um obstáculo à interpretação destas relações de causa e efeito.

O terceiro fator a ser considerado, segundo Andersen e Bro (2010), consiste na presença de valores espúrios (*outliers*), os quais devem ser identificados e, se necessário, removidos. Kettaneh *et al.* (2005) recomendaram a remoção dos valores espúrios através de análise individual de cada variável. Por sua vez, Vinzi *et al.* (2010) afirmam que valores espúrios ‘graves’ devem ser eliminados; contudo, valores espúrios moderados tem um impacto menor sobre o modelo PLS e não precisam ser eliminados. Segundo os mesmos autores, valores espúrios graves são aqueles com desvios superiores a quatro desvios-padrões.

Mehmood *et al.* (2012) apresentam três categorias de métodos de seleção de variáveis, originalmente propostas por Guyon e Elisseeff (2003):

- **Filters** (filtro): a aplicação do método ocorre após o modelo de regressão PLS ser ajustado aos dados. A seleção de variáveis é realizada pela introdução de um filtro aplicado sobre os parâmetros gerados pela PLS, tipicamente representando a relação entre a variável de resposta e a independente (MEHMOOD *et al.*, 2012). Se o parâmetro de uma variável apresentar valor inferior ao limite imposto, ela é descartada; caso contrário, a variável é mantida. Este processo de seleção deve ser realizado somente uma vez após a determinação do modelo de regressão. Uma questão crítica é relacionada ao valor do filtro estabelecido ao parâmetro, que afeta significativamente a seleção; em contrapartida, tais métodos apresentam a vantagem de serem rápidos e de fácil implementação (SAEYS *et al.*, 2007). Exemplos de métodos da categoria *filters* incluem:

a) Método dos coeficientes **b**: baseia-se em calcular um modelo PLS e avaliar cada coeficiente  $b_k$  da regressão PLS em relação a um limite definido pelo pesquisador, sendo que somente as variáveis com coeficientes acima deste limite são mantidas no modelo PLS final (GAUCHI e CHAGNON, 2001; MEHMOOD *et al.*, 2012). Como a operacionalização da regressão PLS apoia-se em variáveis normalizadas, os coeficientes  $b_k$  podem ser utilizados para identificação das variáveis mais relevantes (um baixo valor de  $b_k$  significa que a variável  $k$  tem pouca influência nas variáveis de resposta); de tal forma, todas as variáveis que apresentarem coeficiente  $b_k$  inferior ao limite são eliminadas. Não existe regra consolidada para definição do limite de  $b_k$ ; Liu *et al.* (2008) analisaram o comprimento de onda em aplicações de espectroscopia e escolheram arbitrariamente o valor de  $b_k = 0,05$ , enquanto que Huang *et al.* (2004), trabalhando com seleção de genes, utilizaram procedimentos de validação (cruzada e *Jackknife*) para determinar o valor limite do coeficiente  $b_k$ . A validação *Jackknife* retira uma das observações e utiliza as demais observações para estimar os parâmetros do modelo PLS e prever a observação que foi retirada (LATTIN *et al.*, 2011).

b) Método dos pesos **w**: calculado de maneira semelhante ao método dos coeficientes **b**, utiliza-se o peso **w** como referência para selecionar as variáveis; se o valor observado for superior ao limite definido, a variável permanece; caso contrário, ela é eliminada. Os coeficientes de peso são avaliados para cada componente e o número recomendado de componentes a serem retidos é tipicamente estabelecido por validação cruzada (FRANK, 1987; MEHMOOD *et al.*, 2012).

c) Método VIP (*Variable Importance in the Projection* ou Importância da Variável na Projeção): calcula a importância de cada variável de acordo com a variância explicada de cada componente PLS. Novamente, um limite definido pelo pesquisador é aplicado (geralmente maior do que um) para selecionar as variáveis (CHONG e JUN, 2005). Maiores detalhes acerca do VIP podem ser encontrados em Lazraq *et al.* (2003), Chong e Jun (2005) e Mehmood *et al.* (2012).

- *Wrappers*: os métodos desta categoria integram os métodos de filtro a um algoritmo que extrai subconjuntos de variáveis potencialmente relevantes e avalia cada subconjunto através do ajuste de um modelo PLS às variáveis deste subconjunto (MEHMOOD *et al.*, 2012). Apresentam como desvantagens o alto esforço computacional (em comparação aos métodos de filtro) e o risco de sobreajuste (SAEYS *et al.*, 2007). Como métodos desta categoria destacam-se:

a) Método UVE-PLS (*Uninformative Variable Elimination* ou eliminação de variáveis não informativas): neste método adicionam-se variáveis aleatórias artificiais aos dados previamente à construção da regressão PLS; nesta proposição, as variáveis coletadas devem apresentar importância maior do que as artificiais para permanecerem no modelo (MEHMOOD *et al.*, 2012). O número de variáveis artificiais incluídas tipicamente é o mesmo das variáveis preditoras originais, sendo inseridas no modelo com um valor extremamente baixo de forma a não influenciar significativamente os coeficientes de regressão das variáveis originais. Na sequência, um coeficiente  $v$  é calculado para cada variável, dividindo o coeficiente de regressão  $b_k$  pelo desvio-padrão deste coeficiente (obtido por validação *Jacknife*); o ponto de corte é o máximo valor do coeficiente  $v$  dentre as variáveis aleatórias (FORINA *et al.*, 1999).

b) Método BVE-PLS (*Backward Variable Elimination* ou eliminação inversa de variável): o método inicia com um modelo contemplando todas as variáveis e, em seguida, retira-se uma variável por vez do modelo; avalia-se então o impacto que a retirada de cada uma das variáveis produz no erro de predição como o RMSE (*root mean square error* ou raiz do erro quadrático médio). O procedimento é reproduzido sucessivamente, eliminando uma a uma as variáveis do modelo, até que reste somente uma (Pierna *et al.*, 2009; Mehmood *et al.*, 2012); o subconjunto selecionado é aquele responsável pelo menor RMSE.

O erro de predição RMSE é determinado conforme a equação (8):

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (8)$$

onde  $y_i$  é o valor da variável para a  $i$ -ésima observação;  $\hat{y}_i$  é o valor predito para  $y_i$  e  $n$  é o número de observações.

c) Método IPW-PLS (*Iterative Predictor Weighting* ou preditor ponderado iterativo): proposto por Forina *et al.* (1999), a modelagem via PLS é iterativamente realizada e, a cada ciclo, calcula-se uma medida de importância multiplicando as variáveis predictoras ( $\mathbf{X}$ ) pelos pesos ( $\mathbf{w}$ ). A variável com a menor medida de importância deve ser eliminada e uma nova iteração é realizada através da multiplicação da importância calculada pelos preditores (MEHMOOD *et al.*, 2012).

d) Método BQ (Backward- $Q^2_{\text{cum}}$ ): proposto por Gauchi e Chagnon (2001), gera inicialmente um modelo PLS com todas as variáveis independentes; após a predição das variáveis de resposta, computa-se o critério de seleção  $Q^2_{\text{cum}}$ .  $Q^2_{\text{cum}}$  avalia a capacidade preditiva do modelo na escala de 0 a 1, quanto mais próximo de 1, maior a precisão da predição (GAUCHI e CHAGNON, 2001). É calculado conforme a equação (9).

$$Q^2_{\text{cum}} = 1 - \prod_{j=1}^A \frac{\sum_{i=1}^{n_L} (y_i - \hat{y}_{(-i)})^2}{\sum_{i=1}^{n_L} (y_i - \hat{y}_i)^2} \quad (9)$$

onde  $A$  é o número de componentes retidos PLS na validação cruzada;  $n_L$  é o número de observações no conjunto de treino;  $\hat{y}_{(-i)}$  é o valor previsto para  $y_i$ , enquanto  $y_i$  é retirado dos dados quando da construção do modelo; e  $\hat{y}_i$  é o valor previsto para  $y_i$  enquanto ele é mantido nos dados ao construir o modelo.

Na sequência do método BQ, a variável independente com o menor valor absoluto para o coeficiente de regressão  $b_{mk}$  é eliminada e um novo modelo é calculado para as variáveis restantes; em seguida,  $Q^2_{\text{cum}}$  é recalculado. Tal procedimento de eliminação é repetido sistematicamente até que restar apenas uma variável, e um gráfico relacionando número de variáveis retidas e  $Q^2_{\text{cum}}$  é gerado; o subconjunto de variáveis a ser retido é aquele responsável pelo máximo  $Q^2_{\text{cum}}$ .

Gauchi e Chagnon (2001) compararam métodos de seleção no contexto de PLS em diversos bancos de dados industriais; os autores concluíram que o método BQ é o mais adequado por sua eficiência e simplicidade. Como segunda opção, os autores recomendam substituir o critério  $Q^2_{\text{cum}}$  por RMSE; tal método, denominado BSDEP, alcançou resultados ligeiramente inferiores ao BQ nos dados avaliados.

A literatura aponta ainda outros métodos pertencentes à categoria *Wrappers*, incluindo o CovProc-PLS (*Covariance Procedure* ou procedimento de covariância, IPLS (*Interval-PLS*) e GA-PLS (*Genetic Algorithm* ou algoritmo genético); esses métodos não serão abordados neste trabalho, sendo detalhes a seus respeitos encontrados em Nørgaard *et al.* (2000), Leardi *et al.* (2002) e Reinikainen e Höskuldsson (2007).

- ***Embedded***: nesta categoria, a modelagem e a seleção de variáveis estão incluídas em etapa única. A seleção de variáveis é realizada após a determinação de cada componente do modelo PLS e não apenas ao final da determinação de todos componentes, como ocorre nos métodos acima descritos (MEHMOOD *et al.*, 2012; ABRAHAMSSON *et al.*, 2013). Métodos pertencentes a esta categoria incluem IVS (*Interactive Variable Selection*), PPLS (*Powered PLS*) e SPLS (*Sparse-PLS*); detalhes adicionais sobre tais métodos podem ser encontrados em Mehmood *et al.* (2012).

Apesar do elevado número de abordagens para seleção de variáveis oferecido pela literatura, Mehmood *et al.* (2012) argumentam que não é possível eleger de forma unânime o melhor método de seleção, uma vez que o desempenho dos métodos depende das propriedades dos dados analisados.

A grande maioria das abordagens sobre seleção de variáveis em PLS trata de situações com somente uma variável de resposta, denominado modelo PLS1. Mehmood *et al.* (2012) afirmam que, idealmente, abordagens de seleção no contexto PLS1 devem ser suficientemente simples para serem estendidos ao problema de múltiplas variáveis de resposta (PLS2), o qual é comum em processos industriais (mas que tem sido pouco abordado pela literatura). Alguns autores sugerem utilizar métodos que tratam uma variável de resposta por vez, mas é necessário assumir que essas variáveis de resposta sejam não correlacionadas (o que não é verificado em situações nas quais se deseja encontrar as condições recomendadas de operação para processos industriais em múltiplos estágios).

A literatura em seleção de variáveis no contexto PLS2 é bastante incipiente. Com vistas à determinação dos parâmetros de qualidade relevantes em produtos derivados de tomate, Pedro e Ferreira (2007) compararam a aplicação de modelos PLS1 e PLS2; os autores concluíram que o modelo PLS2 apresentou maior capacidade preditiva do que a maioria dos modelos PLS1, além de reter menor número de componentes do que o PLS1. Já Vandeginste *et al.* (1998) consideraram a habilidade preditiva similar em ambos os modelos e salientaram que é mais vantajoso gerar um único modelo PLS2 do que vários modelos PLS1, principalmente se o número de variáveis de resposta for elevado. Para Ferrer *et al.* (2008), a

capacidade preditiva está associada à correlação existente entre as variáveis de resposta, de maneira que quanto maior esta correlação, melhor tende a ser o desempenho dos modelos PLS2 frente ao PLS1.

Em modelos PLS2, em função do número de variáveis de resposta, pode-se utilizar o erro médio de predição agrupado ( $RMSE_p$ ) ou o coeficiente de correlação agrupado; Pedro e Ferreira (2007) encontraram resultados similares com estes dois critérios. O erro de predição  $RMSE_p$ , para várias variáveis de resposta, pode ser calculado conforme a equação (10):

$$RMSE_p = \sqrt{\frac{\sum_{j=1}^k \sum_{i=1}^{n_t} (y_{ij} - \hat{y}_{ij})^2}{kn_t}} \quad (10)$$

onde  $y_{ij}$  é o valor observado da  $j$ -ésima variável para a  $i$ -ésima observação;  $\hat{y}_{ij}$  é o valor predito para  $y_{ij}$ ;  $k$  é o número de variáveis de resposta e  $n_t$  é o número de observações no conjunto de teste.

## 2.6 PROGRAMAÇÃO QUADRÁTICA

Na análise de sistemas pode-se utilizar uma medida de desempenho global que represente um objetivo a ser alcançado. Em problemas de otimização se busca maximizar ou minimizar esta medida de desempenho global, que depende de diversas outras variáveis, denominadas variáveis de decisão, as quais podem, entre si, serem independentes ou então relacionadas por meio de restrições. Um problema de programação matemática pode ser definido como uma formulação de otimização onde se utilizam funções matemáticas e relações funcionais para descrever os relacionamentos entre as variáveis de decisão (NOCEDAL e WRIGHT, 2006; HILLIER e LIEBERMANN, 2013). Essas relações funcionais definem os tipos de restrições:

- a) de igualdade, na forma de equações (=);
- b) de desigualdade do tipo maior ou igual a ( $\geq$ ) e
- c) de desigualdade do tipo menor ou igual a ( $\leq$ ).

Um modelo geral para o problema de otimização, apresentado por Nocedal e Wright (2006) é apresentado na equação (11):

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{sujeito à} \quad c_i x = 0, i \in \varepsilon \quad e \quad c_i x \geq 0, i \in \iota \quad (11)$$

onde  $f(x)$  é a função objetivo,  $c_i$  são os coeficientes,  $x$  as variáveis de decisão,  $\varepsilon$  o índice para as restrições de igualdade e  $\iota$  o índice para as restrições de desigualdades.

Na programação linear, uma das técnicas de otimização mais utilizadas, a principal característica é que a função objetivo e todas as restrições impostas são representadas por funções lineares (HILLIER e LIEBERMANN, 2013).

Diversas aplicações práticas, no entanto, demandam a adoção de programação não linear para sua adequada modelagem, onde a função objetivo ou as restrições (ou ambas) podem utilizar funções não lineares. Na programação quadrática, as restrições são lineares e somente a função objetivo apresenta a forma quadrática, envolvendo o quadrado de uma variável de decisão (ou mais do que uma) ou o produto de duas variáveis (LACHTERMACHER, 2006; HILLIER e LIEBERMANN, 2013).

O modelo de programação quadrática pode ser representado pela equação (12) para a função objetivo e equação (13) para as restrições (DOSTÁL, 2009).

$$\min_x f(x) = \mathbf{l}x + \frac{1}{2}x'\mathbf{H}x \quad (12)$$

Sujeito às restrições:

$$\mathbf{O}x \leq \mathbf{z} \quad (13)$$

$$\mathbf{S}x = \mathbf{d} \text{ e}$$

$$x \geq 0$$

onde  $\mathbf{l}$  é um vetor linha de termos lineares de  $x$ ;  $\mathbf{H}$  é uma matriz simétrica;  $\mathbf{O}$  é uma matriz com os coeficientes das restrições de desigualdade do tipo ‘menor ou igual’;  $\mathbf{z}$  é um vetor coluna com os valores do lado direito da equação para as restrições de desigualdade;  $\mathbf{S}$  é uma matriz com os coeficientes das restrições de igualdade e  $\mathbf{d}$  é um vetor coluna com os valores do lado direito da equação para as restrições de igualdade.

A condição para se encontrar o mínimo global, nos casos de minimização é que a função objetivo seja uma função convexa, uma vez que as restrições são lineares; no caso de maximização, a função objetivo deve ser côncava, para se encontrar o máximo global (LACHTERMACHER, 2006). Para se identificar se a função objetiva é convexa, é necessário verificar se a matriz  $\mathbf{H}$ , da equação (12) é uma matriz semidefinida positiva (HILLIER e LIEBERMANN, 2013).

A programação quadrática é aplicada nos mais diversos setores, tais como seleção de carteiras de investimento, maximização da margem de lucro, sistemas de abastecimento e demanda, análise de média e variância (McCARL *et al.*, 1977; COLIN, 2011; HILLIER e LIEBERMANN, 2013). A aplicabilidade do método, no entanto, vai além, visto que pode ser utilizado como base para resolução de uma vasta gama de problemas não lineares, através de métodos como Programação Quadrática Sequencial, o qual, conforme Gill e Wong (2012) “encontra uma solução aproximada de uma sequência de subproblemas de programação quadrática em que um modelo quadrático da função objetivo é minimizado sujeito às restrições lineares”. Como exemplo de aplicação da programação quadrática sequencial na determinação de condições ótimas de trabalho, têm-se os trabalhos de Yacoub e MacGregor (2004) e Muteki (2006).

Nocedal e Wright (2006) afirmam que a escolha dos métodos de solução dos problemas de programação quadrática com múltiplas variáveis depende da existência de restrições no modelo. Se o modelo de programação quadrática apresenta somente restrições de igualdade, o método do gradiente conjugado projetado pode ser utilizado para encontrar a solução global; entretanto, quando o problema de programação quadrática convexa apresenta restrições de igualdade e desigualdade, há diversos métodos disponíveis (NOCEDAL e WRIGHT, 2006):

a) Método de conjunto ativo (*active-set method*): de acordo com Luenberger e Ye (2008) o conjunto de restrições é dividido em dois grupos, um tratado como ativo e o outro como inativo, o qual é ignorado. Em cada etapa de um algoritmo o conjunto ativo de restrições no ponto atual, chamado de conjunto de trabalho, é viável e o algoritmo prossegue para encontrar outra solução viável melhorada dentro do conjunto de trabalho; neste ponto, o conjunto de trabalho pode ser alterado.

Este método pode ser aplicado a problemas convexos e não-convexos, sendo eficiente para problemas de tamanho pequeno ou médio. O método Simplex, utilizado em programação linear, é um método de conjunto ativo: partindo de uma solução básica inicial viável, busca, através de multiplicadores de Lagrange e gradientes, a identificação de uma variável básica que deve ser substituída por uma variável não básica, procedimento utilizado até encontrar a solução ótima; na programação quadrática este método difere do Simplex, pois a solução de uma iteração não precisa necessariamente estar no vértice do espaço das soluções viáveis (NOCEDAL e WRIGHT, 2006);

b) Método de projeção de gradiente (*gradient projection method*): este método pertence à classe dos métodos ativos e é eficiente quando as restrições do modelo são lineares, como os limites impostos às variáveis (NOCEDAL e WRIGHT, 2006). Segundo Colin (2011), para “encontrar a direção viável, o método usa a projeção do gradiente da função objetivo nas restrições que estão ativas”; e

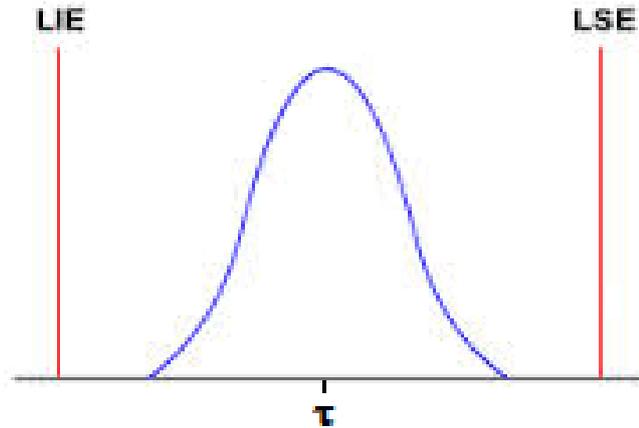
c) Método de ponto interior (*interior-point methods*): este método tem se mostrado eficiente para problemas de grande tamanho. Difere do algoritmo Simplex pelo tipo de busca da solução, pois enquanto o primeiro encontra uma solução básica viável nas bordas ou vértices do espaço de soluções possíveis, o de ponto interior se move na parte interna deste espaço de soluções em direção ao ponto ótimo (procedimento que se mostra mais eficiente quando o número de vértices e restrições é grande. Entretanto, os dois são algoritmos iterativos e apresentam a mesma lógica: (i) encontrar uma solução básica inicial, (ii) a cada iteração encontrar outra solução melhor dentro do espaço das soluções possíveis, e (iii) testar se a solução é ótima, voltando ao passo (ii) caso não seja (NOCEDAL e WRIGHT, 2006; COLIN, 2011; HILLIER e LIEBERMANN, 2013).

## 2.7 CAPACIDADE DO PROCESSO

As características de qualidade desejadas num processo são normalmente expressas sob a forma de um valor nominal. É aceita, no entanto, certa variabilidade neste valor, definida na forma de limites de especificação (MONTGOMERY, 2004). Assim, cada variável de resposta deverá ser caracterizada por um valor nominal ( $\tau$ ), que é o alvo (desejável) para esta variável, bem como por um limite inferior de especificação (LIE) e um limite superior de especificação (LSE), conforme a Figura 7.

Na gestão tradicional de controle de processo através de cartas de controle são utilizados os ‘limites de controle’ baseados nos dados do processo. Os limites de controle para variáveis quantitativas representam a variabilidade natural do processo para a distribuição amostral da média, onde a tolerância (de três desvios-padrões) é subtraída/adicionada à média. São calculados diferentemente de acordo com a carta adotada.

Figura 7 - Limites de especificação de uma característica de qualidade de um produto



Fonte: Adaptada de Siqueira (1997)

Os limites de especificação, quando comparados aos limites de controle, permitem avaliar a capacidade do processo em atender a esta especificação sob a premissa de um processo em controle. A capacidade do processo pode ser medida de várias maneiras, mas dois índices são comumente usados,  $C_p$  e  $C_{pk}$  (MONTGOMERY, 2004). O índice  $C_p$  compara diretamente os limites de especificação e os de controle, conforme a equação (14).

$$C_p = \frac{LSE - LIE}{6\sigma} \quad (14)$$

onde  $LSE$  é o limite superior de especificação;  $LIE$  é o limite inferior de especificação e  $\sigma$  é o desvio-padrão amostral.

Para Siqueira (1997), a avaliação do índice  $C_p$  pode ser resumidamente descrita por: (i)  $C_p < 1$ : o processo não é capaz de atender às especificações; (ii)  $1 \leq C_p \leq 1,33$ : a capacidade do processo em atender às especificações está no limite aceitável; e (iii)  $C_p > 1,33$ : o processo é capaz de atender às especificações.

O índice  $C_p$  é suficiente quando a média do processo coincide com o valor nominal, o que tipicamente não é observado em aplicações práticas. Para evitar esta limitação, calcula-se o índice de desempenho  $C_{pk}$ , de acordo com a equação (15):

$$C_{pk} = \min \left( \frac{\bar{x} - LIE}{3\sigma}, \frac{LSE - \bar{x}}{3\sigma} \right) \quad (15)$$

Valores idênticos de  $C_{pk}$  e  $C_p$  determinam que o processo está centrado; caso contrário, o processo está fora de centro (valor nominal diferente da média do processo). Combinações de diferentes valores de  $C_p$  e  $C_{pk}$  permitem interpretar a capacidade do processo, a necessidade

de ajuste no valor médio ou na variabilidade do processo, e na alteração dos limites de especificação (SIQUEIRA, 1997).

Neste trabalho, os índices de capacidade são utilizados como indicadores da capacidade de atendimento às especificações de acordo com as amostras obtidas.

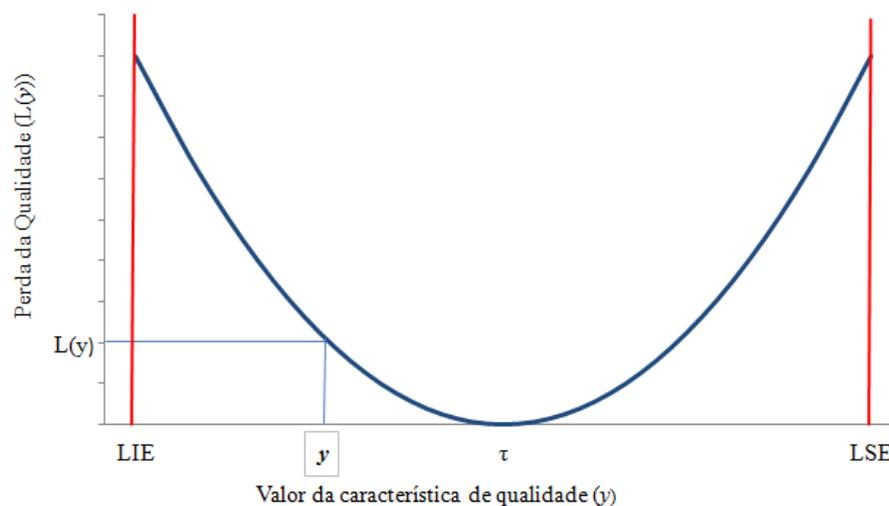
## 2.8 FUNÇÃO PERDA DE QUALIDADE DE TAGUCHI

A premissa da função perda de qualidade de Taguchi é que qualquer desvio em relação a um valor nominal de uma característica de qualidade é considerado uma perda ao consumidor; deve-se então minimizar esses desvios, produzindo produtos de melhor qualidade (OAKLAND, 2008). A função perda tem sido utilizada na avaliação de processos, analisando a relação entre a capacidade do processo e a qualidade do produto, permitindo determinar o impacto monetário de desvios do valor alvo (SIQUEIRA, 1997). A função perda de qualidade de Taguchi é apresentada na equação (16):

$$L(y) = \frac{CT}{\Delta^2} (y - \tau)^2 \quad (16)$$

onde  $L(y)$  é a perda de qualidade, avaliada financeiramente em relação ao desvio do valor nominal;  $CT$  é o custo de produção do produto;  $\Delta$  é a tolerância em torno do valor nominal,  $y$  é o valor obtido para a variável de resposta e  $\tau$  é o valor nominal da variável de resposta. A Figura 8 ilustra a função perda de qualidade de Taguchi.

Figura 8 - Função perda de qualidade de Taguchi



Fonte: Adaptada de Siqueira (1997)

### 3 METODOLOGIA

Neste capítulo é apresentado o método proposto para a seleção de variáveis independentes e determinação da condição ótima de trabalho em processos contínuos multivariados e em múltiplos estágios. O método apoia-se nos seguintes passos: concluída a coleta de dados e pré-tratamento dos mesmos, geram-se modelos de regressão PLS para cada estágio (definindo o número de componentes por validação cruzada); posteriormente, selecionam-se as variáveis mais relevantes em cada modelo. Em seguida, levantam-se os limites de especificação das variáveis de resposta e gera-se um modelo de programação quadrática com vistas à obtenção da condição ótima de operação.

No que tange as sistemáticas de seleção de variáveis, o método proposto integra de forma original as proposições BQ de Gauchi e Chagnon (2001) e BVE-PLS de Pierna *et al.* (2009). Para tanto, utiliza o procedimento de validação cruzada sugerido pelo método BQ, bem como os coeficientes de regressão PLS para ordenar as variáveis de acordo com sua importância. Ao método BVE-PLS deve-se o critério de seleção RMSE e o procedimento gráfico de determinação do número de variáveis retidas (que relaciona RMSE e número de variáveis retidas). Além disto, o método aqui proposto inova ao estender sua aplicação a processos de múltiplos estágios e com múltiplas variáveis dependentes, visto que as proposições acima mencionadas se restringem a problemas com uma única variável dependente e único estágio.

#### 3.1 COLETA E PRÉ-TRATAMENTO DOS DADOS

Dados podem ser definidos como observações documentadas de itens, geralmente por meio de números, provenientes de medidas ou contagens. Diferentes características dos itens são descritas por variáveis (STEVENSON, 1981). Processos produtivos industriais tipicamente são descritos por um elevado número de variáveis, as quais incluem tempos, temperaturas, pesos, comprimentos, áreas, fluxos, velocidades e pressões, entre outras.

Tais informações podem ser coletadas de forma automática ou manual. Em processos industriais, sensores mecânicos ou eletrônicos coletam automaticamente e registram dados obtidos em equipamentos presentes no processo (balanças, termômetros, cronômetros,

contadores, etc.) ou no ambiente (termômetros, barômetros, anemômetros, etc). Outros dados podem depender da observação humana com registro manual em instrumento de coleta (planilha ou gráfico).

Em processos contínuos em múltiplos estágios, cada estágio deve ser identificado e ter seus limites (início e fim) determinados, considerando a abordagem de conhecimento científico, onde se assume alteração das características de qualidade do produto. O processo gerador dos dados deve ser avaliado, permitindo localizar fisicamente o ponto de coleta de cada variável independente e de resposta e possibilitando a associação variável-estágio.

Após a obtenção dos dados, recomenda-se executar os seguintes procedimentos: (i) classificar cada variável coletada como sendo de natureza quantitativa ou qualitativa (variáveis quantitativas são preferenciais em modelos de regressão PLS); (ii) identificar o estágio a que cada variável (independentes e de resposta) pertence; (iii) avaliar a presença e descarte de valores espúrios (*outliers*), e (iv) realizar a normalização dos dados. A normalização consiste em um pré-processamento recomendado pelo algoritmo NIPALS, de forma a colocar as variáveis em uma escala semelhante através da equação (17). A normalização também permite utilizar o coeficiente de regressão PLS como indicador da relevância das variáveis independentes na explicação das variáveis de resposta.

$$Z_i = (x_i - \mu) / \sigma \quad (17)$$

onde  $Z_i$  é o valor normalizado da  $i$ -ésima observação,  $x_i$  é o valor observado da  $i$ -ésima observação;  $\mu$  é a média da população para a variável e  $\sigma$  é o desvio-padrão da variável em análise.

Como os dados são uma amostra da população (portanto apresentam distribuição amostral  $t$  de Student), os parâmetros da média e desvio-padrão populacionais são estimados pelas médias amostrais; para se aproximar da distribuição normal, é preciso garantir que o número de observações seja superior a quarenta, de acordo com o Teorema do Limite Central (STEVENSON, 1981).

A identificação de valores espúrios deve ser realizada através de análise individual de cada variável, de acordo com Kettaneh *et al.* (2005), seguindo da remoção dos valores espúrios acima de quatro desvios-padrões.

### 3.2 GERAÇÃO DE UM MODELO DE REGRESSÃO PLS PARA CADA ESTÁGIO

A modelagem proposta baseia-se na representação de processos em cascata de múltiplos estágios, proposta por Jearkpaporn *et al.* (2007) e ilustrada na Figura 4. Os estágios são numerados sequencialmente na ordem de execução das tarefas e representados por  $q$  ( $q = 1, \dots, Q$ ). As variáveis de resposta, que apresentam as características de qualidade do produto, são representadas por  $y$ , podendo ser observadas ao final de cada estágio. O número de variáveis de resposta pode variar em cada estágio ( $m = 1, \dots, M$ ), sendo  $y_{qm}$  uma variável de resposta  $m$  no estágio  $q$ .

Da mesma forma, as variáveis que descrevem os parâmetros do processo são as variáveis independentes, denominadas por  $x$ , observáveis nos estágios do processo. O número de variáveis independentes também pode variar em cada estágio ( $k = 1, \dots, K$ ); de tal forma,  $x_{qk}$  representa uma variável independente  $k$  no estágio  $q$ .

Respeitando o efeito cascata descrito por Hawkins (1993), uma variável  $y_{q'm}$ , observada em um estágio  $q'$ , somente pode ser influenciada por variáveis independentes se o estágio desta variável for igual ou precedente do estágio onde  $y$  é observada, ou seja,  $q \leq q'$ .

Já em relação às variáveis do processo (independentes), não se pode supor independência entre elas, pois podem estar correlacionadas (por exemplo, a pressão de vapor e a temperatura num vaso de pressão, num ponto do processo, ou a velocidade da linha e o tempo de processamento). Nesse caso, objetiva-se, com o modelo de regressão PLS a ser gerado, escolher um subconjunto das variáveis independentes que melhor predigam as variáveis de resposta.

A geração das regressões PLS para cada estágio apoia-se nas seguintes premissas: (i) a definição de início e término de cada estágio ocorre analisando-se as alterações das características de qualidade do produto, o que geralmente coincide com a separação física; (ii) o fluxo de processamento é bem determinado durante todo o processo; (iii) uma característica de qualidade do produto pode ser observada em um estágio qualquer (por exemplo: o peso de uma peça pode ser observado no segundo estágio e outra característica, como comprimento, somente no último); (iv) as características de qualidade de cada estágio são conhecidas e observáveis, com limites de especificação determinados; (v) erros de coleta nos valores das variáveis de processo são negligenciáveis; (vi) as variáveis de resposta são analisadas por

estágio, separadamente; e (vii) as variáveis de resposta de um estágio  $q$  podem ser afetadas por todas as variáveis independentes deste estágio ou de seus antecessores ( $x_{qm}=1, \dots, Q$ ).

Desta maneira, em cada estágio, o modelo PLS gerado é composto pelas variáveis de resposta do estágio e por todas as variáveis independentes desde o primeiro estágio até o estágio em análise.

Na sequência, define-se o número de componentes (variáveis latentes) a serem retidos na regressão PLS através do procedimento de validação cruzada (do tipo *n-fold cross validation*). Os dados coletados na etapa 3.1 são então randomicamente separados em cinco grupos (representados por A, B, C, D e E). Quatro destes grupos são inseridos em uma porção de treino (para geração do modelo PLS) e o grupo remanescente é inserido em uma porção de teste (utilizado para validar a capacidade preditiva do modelo PLS gerado). Tal proporção entre treino e teste atende à indicação de Wold *et al.* (2001). Para cada combinação de 4 grupos na porção de treino e 1 na porção de teste, gera-se uma regressão PLS e altera-se o número de variáveis latentes de 1 até 20; por fim, avalia-se o erro de predição  $RMSE_p$  (ver equação 10). O número de componentes retidos responsável pelo menor  $RMSE_p$  é então utilizado nos modelos PLS durante a etapa de seleção de variáveis, descrita na sequência.

### 3.3 SELEÇÃO DAS VARIÁVEIS INDEPENDENTES MAIS RELEVANTES PARA CADA REGRESSÃO PLS EM CADA ESTÁGIO DO PROCESSO

O método proposto para a seleção de variáveis apoia-se na magnitude dos coeficientes  $b_k$  da regressão PLS (ver Equação 6) de cada estágio para mensurar a importância das variáveis independentes para a predição das variáveis de resposta. Um baixo coeficiente  $b_k$  denota não haver relação significativa entre uma variável independente  $x_k$  e a de resposta. A utilização da magnitude é válida para seleção de variáveis, tendo em vista que as variáveis independentes foram normalizadas antes da geração dos modelos PLS. Tal sistemática de seleção foi proposta por Gauchi e Chagnon (2001) e Chong e Jun (2005), mas em ambos o método implementado era do tipo *filters*, com a utilização de um valor de corte como referência na determinação das variáveis independentes importantes. Neste trabalho, o coeficiente  $b_k$  não é associado a um filtro, mas a uma eliminação sistemática das variáveis do tipo *backwards*.

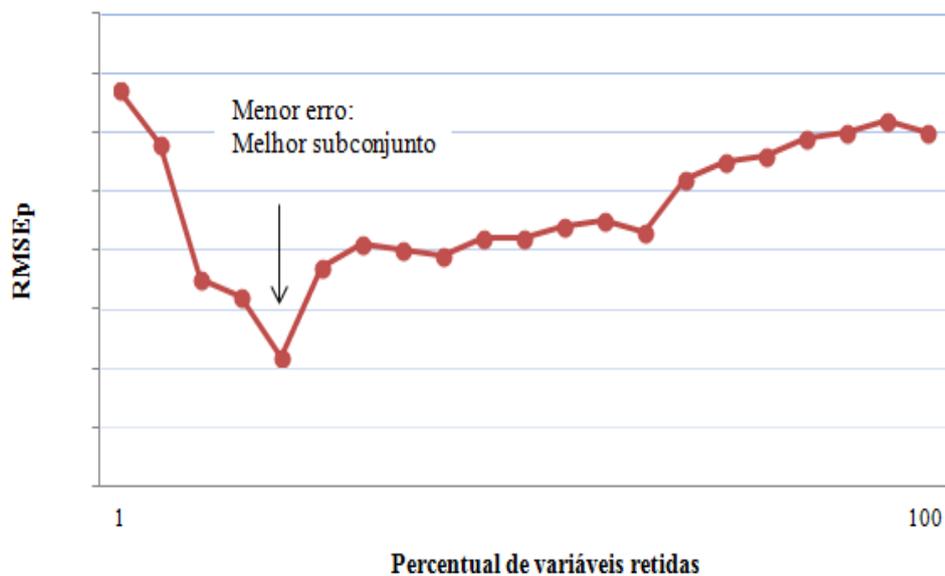
Considere um estágio  $q$  monitorado por múltiplas variáveis de resposta organizadas em uma matriz  $\mathbf{Y}_q$  e independentes em uma matriz  $\mathbf{X}_q$ . Após aplicar-se a regressão PLS neste estágio, gera-se uma matriz  $\mathbf{B}_q$  composta por elementos  $b_{qmk}$ , que associam as  $m$  variáveis de resposta às  $k$  variáveis independentes no estágio  $q$ . Na sequência, gera-se um indicador de importância da variável independente  $k$  através da equação (18).

$$R_{qk} = \sum_{m=1}^M |b_{qmk}| \quad \text{para } q=1, \dots, Q \quad (18)$$

onde  $R_{qk}$  quantifica a importância da variável  $k$  no estágio  $q$  levando em consideração todas as variáveis de resposta afetadas pela variável independente  $k$ .

Para cada estágio  $q$ , gera-se inicialmente uma regressão PLS utilizando as  $K$  variáveis independentes e calcula-se o  $\text{RMSE}_p$ . Na sequência, as variáveis independentes são ordenadas decrescentemente de acordo com o índice de importância  $R_{qk}$  e a variável com o menor  $R_{qk}$  é eliminada. Um novo modelo PLS para este estágio é então construído com base nas  $K-1$  variáveis remanescentes e o erro de previsão  $\text{RMSE}_p$  é novamente calculado. O procedimento iterativo proposto é repetido até restar apenas uma variável. Recomenda-se então construir um gráfico associando erro de previsão ( $\text{RMSE}_p$ ) e número de variáveis retidas; o subconjunto de variáveis responsável pelo menor  $\text{RMSE}_p$ , é selecionado para o estágio  $q$ , conforme ilustrado no perfil genérico da Figura 9. No caso de valores idênticos de  $\text{RMSE}_p$  serem obtidos durante o processo de remoção das variáveis, recomenda-se optar pelo subconjunto que retiver menor número de variáveis.

Figura 9 – Perfil do erro de previsão em função do número de variáveis retidas



Fonte: Adaptada de Zimmer e Anzanello (2013)

O procedimento é repetido para os  $Q$  estágios analisados, gerando então  $Q$  regressões PLS (uma para cada estágio) com variáveis selecionadas com base em sua capacidade preditiva. Alternativamente, pode-se substituir  $RMSE_p$  por  $Q^2_{cum}$  como métrica de qualidade de predição, de forma semelhante ao proposto em Gauchi e Chagnon (2001).

Por fim, as  $Q$  regressões PLS geradas a partir das variáveis selecionadas são utilizadas na predição das variáveis dependentes de cada estágio, as quais são então inseridas na formulação de otimização com vistas à definição das condições ótimas de operação do processo. Tal procedimento é detalhado no próximo passo.

### 3.4 FORMULAÇÃO DO MODELO DE PROGRAMAÇÃO QUADRÁTICA

As regressões PLS geradas em 3.3 (compostas pelas variáveis independentes selecionadas em cada estágio) são então inseridas em uma formulação de otimização com vistas à minimização dos valores preditos em relação às especificações das variáveis de resposta. Para tanto, deve-se levantar os valores nominais e limites de especificação das variáveis de resposta de todos os estágios. Caso alguma variável não conte com valores estabelecidos para tais limites, pode-se coletar a opinião de especialistas para a determinação destes valores ou exclusão da variável. A opinião de especialistas também pode ser utilizada para quantificar subjetivamente a importância das características de qualidade; tal quantificação pode então ser utilizada como um fator ponderador de importância no modelo de otimização, priorizando variáveis de resposta tidas como mais relevantes para o processo em questão. Para tanto, sugere-se a geração de um índice  $I_{qm}$ , onde 1 representa uma variável de resposta  $m$  com importância normal no estágio  $q$ , ao passo que valores superiores a 1 sinalizam variáveis de resposta (descritoras de características de qualidade do produto) mais relevantes. As informações coletadas podem ser armazenadas conforme o arranjo do Quadro 1.

No presente trabalho, não se pretende avaliar a perda financeira produzida pelos desvios (foco tradicional da Função de Taguchi), mas reduzir esses desvios com vistas ao atendimento das características de qualidade nominais. Para tanto, substitui-se o custo total de fabricação (termo  $CT$ , na equação (16)) pelo indicador  $I_{qm}$ , que representa o índice de importância da característica de qualidade, resultando na equação (19). Os demais termos seguem as definições da equação (16).

Quadro 1 - Limites de especificação, valor nominal e índice de importância

Variável	Limite Inferior de Especificação (LIE)	Valor Nominal ( $\tau$ )	Limite Superior de Especificação (LSE)	Índice Importância ( $I_{qm}$ )
1				
2				
...				
$m$				
...				
$M$				

Fonte: Elaborada pelo autor

$$D(y) = \frac{I_{qm}}{\Delta^2} (y - \tau)^2 \quad (19)$$

Na sequência, gera-se um modelo de programação quadrática que visa minimizar o desvio entre os valores previsto e alvo de todas as variáveis, em todos os estágios, utilizando a função em (19). A tolerância  $\Delta$  é representada em termos dos limites de especificação da variável de resposta, de acordo com a equação (20).

$$\Delta = \frac{LSE_{qm} - LIE_{qm}}{2} \quad (20)$$

No caso de haver somente um limite de especificação inferior, a tolerância é determinada pela diferença entre o valor nominal para a variável em questão e o limite de especificação inferior, de acordo com a equação (21).

$$\Delta = \tau_{qm} - LIE_{qm} \quad (21)$$

onde  $\tau_{qm}$  é o valor nominal para a variável  $m$  no estágio  $q$  e  $LIE_{qm}$  é o limite inferior de especificação para a variável  $m$  no estágio  $q$ .

De maneira semelhante, se apenas o limite superior de especificação for estabelecido, pode-se determinar a tolerância como a diferença entre o limite superior de especificação e o valor nominal da variável, conforme a equação (22).

$$\Delta = LSE_{qm} - \tau_{qm} \quad (22)$$

onde  $LSE_{qm}$  é o limite superior de especificação para a variável  $m$  no estágio  $q$  e  $\tau_{qm}$  é o valor nominal para a variável  $m$  no estágio  $q$ .

Representando estas alterações na função objetivo, obtém-se a equação (23).

$$\text{Min } Z = \sum_{q=1}^Q \sum_{m=1}^M D_{qm} = \sum_{q=1}^Q \sum_{m=1}^M I_{qm} \cdot \frac{(\hat{y}_{qm} - \tau_{qm})^2}{(\Delta)^2} \quad (23)$$

onde  $D_{qm}$  é o desvio do valor nominal da variável  $m$  no estágio  $q$ ;  $I_{qm}$  é o índice de importância da variável  $m$  no estágio  $q$ ;  $\hat{y}_{qm}$  é o valor previsto para a variável  $m$  no estágio  $q$ ;  $\tau_{qm}$  é o valor nominal para a variável  $m$  no estágio  $q$ ;  $\text{LSE}_{qm}$  é o limite superior de especificação para a variável  $m$  no estágio  $q$  e  $\text{LIE}_{qm}$  é o limite inferior de especificação para a variável  $m$  no estágio  $q$ .

A utilização da forma quadrática no numerador e no denominador garante que, independente da magnitude da variável analisada, o desvio seja medido de forma proporcional à amplitude da especificação.

As restrições do modelo de programação quadrática são formadas:

a) pelas regressões PLS obtidas em cada um dos estágios, descrevendo as relações entre as variáveis independentes retidas e as variáveis de resposta, nas respectivas escalas originais. As regressões, conforme a equação (6), são transformadas em restrições, de acordo com a equação (24):

$$\hat{y}_{qm} - \sum_k b_{mk} x_{qk} = \alpha_{mk} \quad (24)$$

onde  $\hat{y}_{qm}$  é o valor previsto para a variável  $m$  no estágio  $q$ ;  $\alpha_{mk}$  é o valor do termo independente da regressão para a variável de resposta  $m$  e a variável independente  $k$ ;  $b_{mk}$  é o valor do coeficiente de regressão para a variável de resposta  $m$  e a variável independente  $k$  e  $x_{qk}$  é o valor da variável independente  $k$  no estágio  $q$ ; e

b) pelos limites de operação (inferior e superior) das variáveis independentes retidas, obtidos nos dados coletados, de forma que não haja extrapolação no modelo de otimização e garantindo que a solução obtida esteja dentro de condições operacionais já vivenciadas na prática, conforme recomendado por Jaeckle e Macgregor (1998) e Tomba *et al.* (2012).

### 3.5 OTIMIZAÇÃO E ANÁLISE DOS RESULTADOS

Com a formulação do modelo de programação quadrática definida, a otimização é executada através de aplicativo computacional, utilizando um algoritmo de ponto interior. Precede-se então à análise dos resultados gerados pelo modelo, comparando os resultados

obtidos após a seleção de variáveis com os resultados obtidos pela regressão PLS com todas as variáveis independentes no modelo. Esta comparação pode ser feita através de vários indicadores, mas será limitada ao valor do coeficiente de explicação e do erro médio de previsão agrupado  $RMSE_p$ . Os resultados também devem ser coerentes com a realidade do processo, principalmente em relação aos patamares recomendados para as variáveis independentes e para tanto, serão avaliados por um especialista. Por fim, recomenda-se, no método proposto, que as variáveis independentes não selecionadas pelo método sejam avaliadas no primeiro instante de aplicação no processo para corroborar sua irrelevância.

## 4 EXEMPLOS NUMÉRICOS

Neste capítulo, a sistemática proposta é inicialmente aplicada a dados reais provenientes de uma empresa avícola. Na sequência, a sistemática é aplicada em um estudo simulado com baseado em dados de um processo de fabricação industrial de látex. O estudo simulado é justificado pelo reduzido número de variáveis e observações contidas nos dados reais.

### 4.1 ESTUDO DE CASO EM UMA EMPRESA AVÍCOLA

O estudo de caso foi realizado com dados de uma empresa do ramo avícola da região Oeste de Santa Catarina, que solicitou sigilo quanto à divulgação de dados e detalhes do processo. Os dados são provenientes de etapas distintas do processo de produção de carne de perus, incluindo criação de aves, abate, evisceração, espostejamento e corte. Os dados foram coletados de forma manual, o que é comum neste tipo de processo, sendo as variáveis de processo (independentes) obtidas por meio de registro manual, sendo a maioria delas obtidas em equipamentos do processo (balanças, termômetros e contadores).

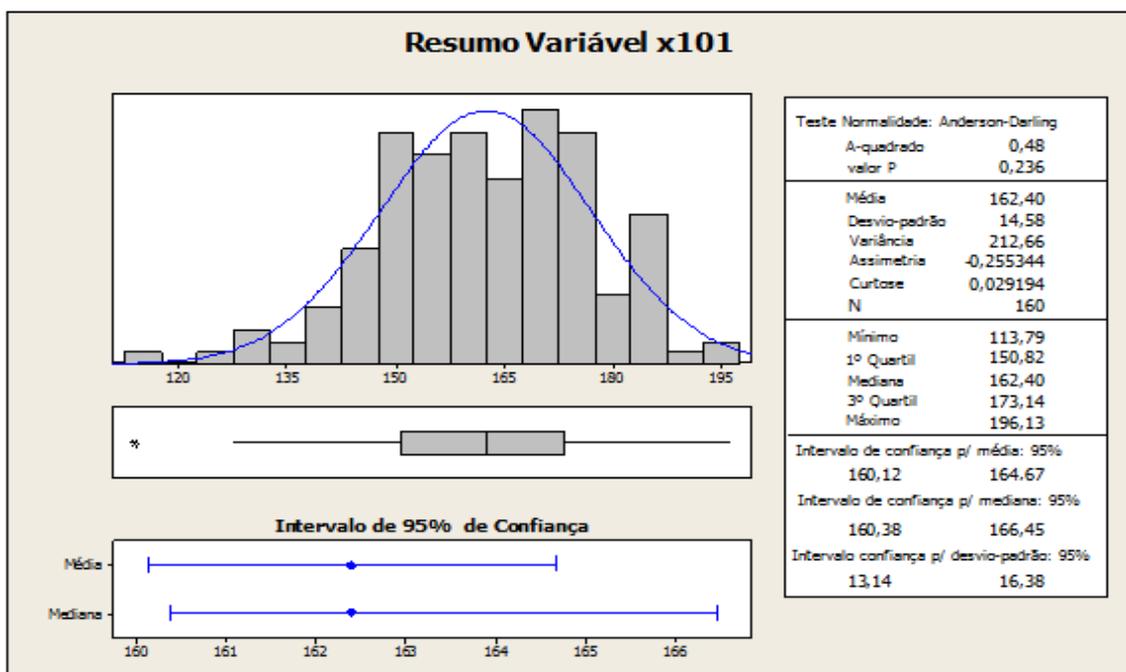
O processo foi dividido em três estágios relativos às etapas de produção: o primeiro estágio compreende a etapa de criação de aves, o segundo as etapas de abate e evisceração, e o terceiro refere-se às etapas de espostejamento e corte. Na Figura 1, observa-se que a etapa de criação de aves é a última do segmento avicultura e na sequência, o processo segue no abatedouro, onde se encontram as etapas seguintes, apresentadas na Figura 2. A separação das etapas em estágios é baseada na abordagem de conhecimento científico e na experiência de especialistas de processo. Os locais de coleta de dados utilizados na empresa e as variáveis coletadas limitaram a separação do processo em mais estágios do que o definido. Assim, como as características de qualidade são avaliadas após a evisceração, agrupou-se num estágio as etapas de abate e evisceração; da mesma maneira, como as variáveis de resposta são avaliadas após o término do espostejamento, as etapas de espostejamento e corte foram agrupadas num estágio.

Na sequência, avaliou-se a base de dados associada aos três estágios criados, junto a um especialista, mantendo-se as variáveis quantitativas de processo e de resposta. Concluída

esta análise inicial das variáveis, obtiveram-se 20 variáveis de processo (independentes) e 22 variáveis de qualidade (resposta). Foram retidas 189 observações após a eliminação de registros com valores faltantes.

As variáveis foram então avaliadas individualmente quanto à presença de valores espúrios através do software Minitab 14.1® (ver exemplo para variável  $x_{101}$  na Figura 10). Observou-se que 15 variáveis (9 independentes e 6 de resposta) das 42 variáveis iniciais apresentam um comportamento assimétrico, o que é esperado para variáveis do tipo de processo analisado. Na sequência, através de análise individual de cada variável, procedeu-se ao descarte de valores extremos (representados por valores acima de quatro desvios-padrões). Ao final da análise preliminar, foram descartadas 29 observações, restando 160 observações no banco a ser analisado.

Figura 10 – Análise das variáveis quanto à presença de valores espúrios



Fonte: Elaborada pelo autor

Em função do número reduzido de variáveis independentes disponíveis (e recomendação subjetiva de especialistas da empresa de que a maioria das variáveis independentes deveriam ser retidas), foram inseridas duas variáveis aleatórias artificiais (com média 0 e desvio-padrão 1) na matriz de variáveis independentes de cada estágio. Tais variáveis são representadas, no primeiro estágio, por  $x_{110}$  e  $x_{111}$ ; no segundo estágio por  $x_{206}$  e  $x_{207}$ , e no terceiro por  $x_{307}$  e  $x_{308}$ . Tal inserção elevou o número de variáveis para 48 (26 independentes e 22 de resposta). O objetivo da inclusão dessas variáveis é avaliar a

capacidade do método de apontar tais variáveis como irrelevantes e indicar o seu descarte. Caso as variáveis artificiais não sejam eliminadas pela seleção, ficam evidenciadas ineficiências no método de seleção. O conjunto final de variáveis independentes é apresentado no Quadro 2.

Quadro 2 - Classificação das variáveis por tipo e estágio

Estágio	Tipo	Identificação	Descrição da variável
1	Independente	$x_{101}$	Tempo de processamento
1	Independente	$x_{102}$	Viabilidade na criação
1	Independente	$x_{103}$	Insumo (ração/ave)
1	Independente	$x_{104}$	Custo no processamento
1	Independente	$x_{105}$	Despesa no processamento
1	Independente	$x_{106}$	Tempo de espera
1	Independente	$x_{107}$	Quantidade alojada
1	Independente	$x_{108}$	Transporte
1	Independente	$x_{109}$	Densidade alojada
1	Independente	$x_{110}$	Artificial 1
1	Independente	$x_{111}$	Artificial 2
1	Resposta	$y_{101}$	Peso médio vivo
1	Resposta	$y_{102}$	Custo marginal na criação
1	Resposta	$y_{103}$	Viabilidade total
1	Resposta	$y_{104}$	Quantidade abatida
1	Resposta	$y_{105}$	Eficiência
1	Resposta	$y_{106}$	Custo (parcela)
2	Independente	$x_{201}$	Tempo de processamento
2	Independente	$x_{202}$	Temperatura (processo)
2	Independente	$x_{203}$	Temperatura (produto)
2	Independente	$x_{204}$	Velocidade da linha
2	Independente	$x_{205}$	Densidade no transporte
2	Independente	$x_{206}$	Artificial 3
2	Independente	$x_{207}$	Artificial 4
2	Resposta	$y_{201}$	Peso (produto)
2	Resposta	$y_{202}$	Custo (carcaça)
2	Resposta	$y_{203}$	Eficiência (parcial)
2	Resposta	$y_{204}$	Eficiência (processo)
2	Resposta	$y_{205}$	Temperatura da carcaça na saída
2	Resposta	$y_{206}$	Peso (subproduto)
3	Independente	$x_{301}$	Peso perna (L)
3	Independente	$x_{302}$	Peso peito (P)
3	Independente	$x_{303}$	Peso pescoço (U)
3	Independente	$x_{304}$	Peso gordura (G)
3	Independente	$x_{305}$	Peso asa (A)
3	Independente	$x_{306}$	Velocidade de espostejamento
3	Independente	$x_{307}$	Artificial 5
3	Independente	$x_{308}$	Artificial 6
3	Resposta	$y_{301}$	Rendimento filé de peito (PF)
3	Resposta	$y_{302}$	Rendimento peito (PI)
3	Resposta	$y_{303}$	Rendimento retalhos (RP)
3	Resposta	$y_{304}$	Rendimento sobrecoxa (S)
3	Resposta	$y_{305}$	Rendimento coxa (C)
3	Resposta	$y_{306}$	Rendimento meio da asa (AM)
3	Resposta	$y_{307}$	Rendimento ponta da asa (AP)
3	Resposta	$y_{308}$	Rendimento pescoço (U)
3	Resposta	$y_{309}$	Custo após processamento
3	Resposta	$y_{310}$	Temperatura P

Fonte: Elaborado pelo autor

Por sua vez, as variáveis de resposta analisadas são caracterizadas por dimensões distintas, descrevendo qualidade intrínseca do produto, custo e eficiência obtida no processamento, dentre outras dimensões.

Para um maior reconhecimento das relações entre variáveis independentes e de resposta, propõe-se avaliar a correlação entre os blocos de variáveis, valendo-se do Quadro 3. Percebe-se que 3 variáveis de resposta do primeiro estágio (de 6 variáveis), 3 variáveis (de 6) do segundo estágio e 5 variáveis do último estágio (de 10) apresentam correlações moderadas ou superiores (ver Apêndice A).

Quadro 3 – Interpretação do coeficiente de correlação

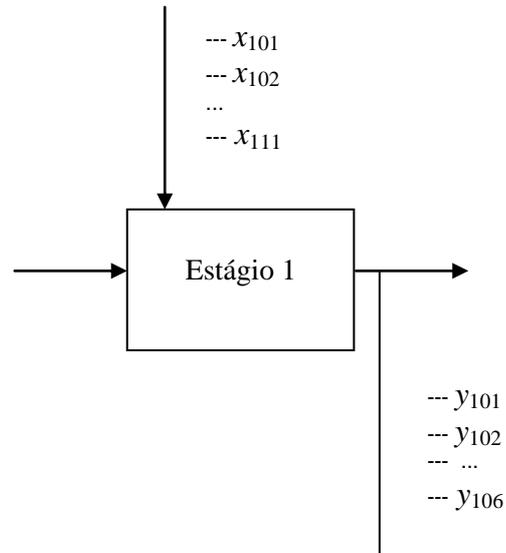
Faixa de r	Intensidade da correlação
0,91 a 1,00	Muito forte
0,71 a 0,90	Alta ou forte
0,41 a 0,70	Moderada
0,21 a 0,40	Pequena, mas definida
0,00 a 0,20	Leve, quase imperceptível

Fonte: Hair, Jr. Josepf F. et al. (2005)

Algumas variáveis independentes também apresentam correlações significativas, sendo 3 variáveis no primeiro estágio, 5 no segundo e 6 no terceiro. Nota-se relacionamento nas variáveis independentes entre estágios: 2 variáveis do segundo estágio têm elevados níveis de correlação com 3 do primeiro, ao passo que 6 variáveis do último estágio estão altamente correlacionadas com 2 variáveis do segundo estágio e com 3 do primeiro estágio (ver Apêndice A). A presença destas correlações, dentro de um estágio e entre estágios, pode trazer dificuldade na identificação da variável independente que provoca a alteração da variável de resposta (relação causa-efeito).

Concluída a separação do processo em estágios e análise preliminar das variáveis e observações, geram-se os modelos de regressão PLS para cada estágio. O modelo de regressão PLS do primeiro estágio partiu de 11 variáveis independentes e 6 variáveis de resposta, conforme a Figura 11.

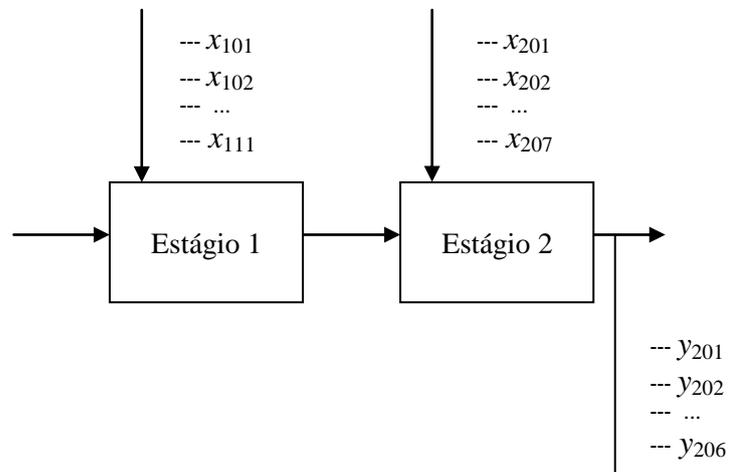
Figura 11 – Representação esquemática das variáveis do estágio 1



Fonte: Elaborada pelo autor

O segundo estágio tem o seu modelo composto por 6 variáveis de resposta e 18 variáveis independentes, sendo 11 do estágio 1 e 7 do estágio 2, conforme a Figura 12.

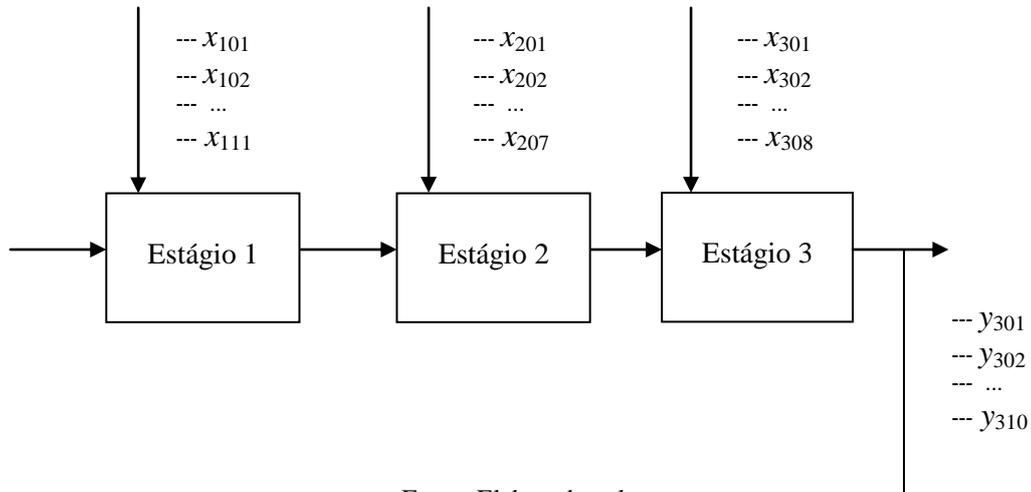
Figura 12 – Representação esquemática das variáveis do estágio 2



Fonte: Elaborada pelo autor

O modelo do estágio 3 é composto por 10 variáveis de resposta e 26 variáveis independentes, das quais 11 são do primeiro estágio, 7 do segundo e 8 do terceiro (ver Figura 13).

Figura 13 – Representação esquemática das variáveis do estágio 3



Fonte: Elaborada pelo autor

Na sequência, dividiu-se o banco de dados em cinco grupos para realização do procedimento de validação cruzada para cada estágio. Em cada grupo, separaram-se as observações em porções de treino (80% das observações) e teste (20% das observações), conforme representação do Quadro 4.

As regressões PLS para cada estágio foram geradas utilizando o algoritmo NIPALS. Ao todo, foram gerados 15 modelos na etapa de validação cruzada (cruzamento de 5 grupos e 3 estágios) utilizando as variáveis de resposta daquele estágio e as independentes daquele estágio acrescidas das independentes dos estágios anteriores. Para gerar estes modelos, utilizou-se o software Matlab<sup>®</sup> (versão 7.12); as rotinas estão disponíveis no Apêndice B. Para cada regressão, gerou-se um gráfico de  $RMSE_p$  versus número de variáveis latentes, buscando-se definir o número de componentes a ser retido em cada modelo.

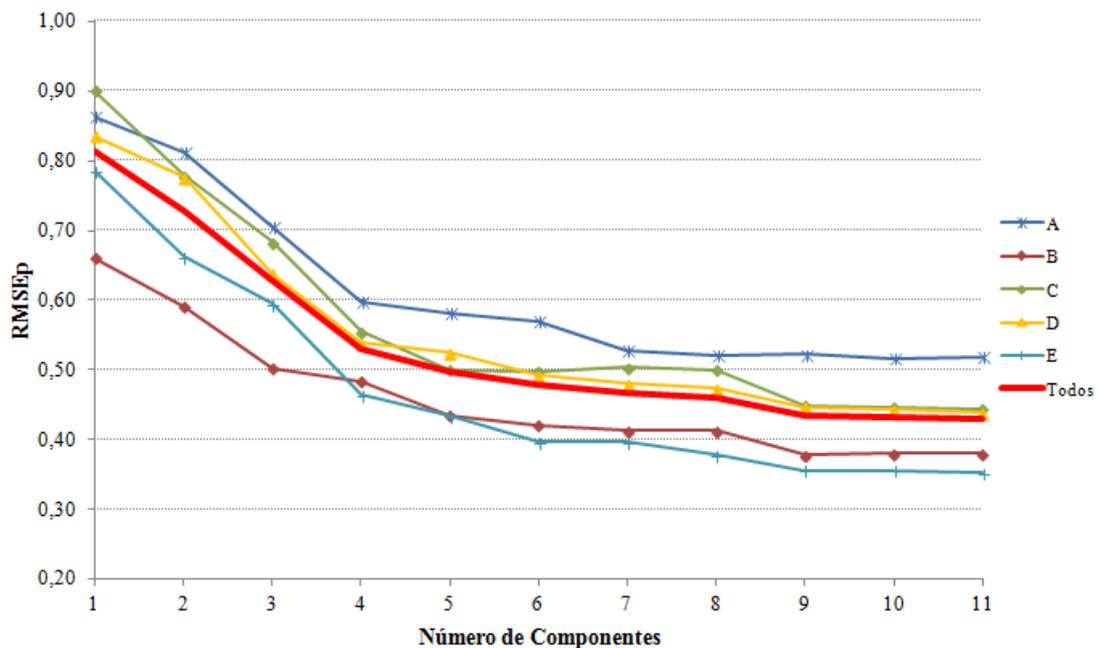
Quadro 4 - Formação dos grupos para a validação cruzada

Observações	Grupo				
	A	B	C	D	E
1 a 32	Treino	<b>Teste</b>	Treino	Treino	Treino
33 a 64	Treino	Treino	<b>Teste</b>	Treino	Treino
65 a 96	Treino	Treino	Treino	<b>Teste</b>	Treino
97 a 128	Treino	Treino	Treino	Treino	<b>Teste</b>
129 a 160	<b>Teste</b>	Treino	Treino	Treino	Treino

Fonte: Elaborado pelo autor

Para o estágio 1, foram retidos 4 componentes, conforme a Figura 14, onde o comportamento do critério  $RMSE_p$  de cada grupo (A a E) pode ser visualizado. A linha mais espessa, denominada ‘Todos’, representa o  $RMSE_p$  de todos os grupos, calculado através da equação (10). A utilização deste cálculo agrupado tem como objetivo auxiliar na determinação do número de componentes para o estágio, uma vez que há variação no comportamento para cada grupo da validação cruzada.

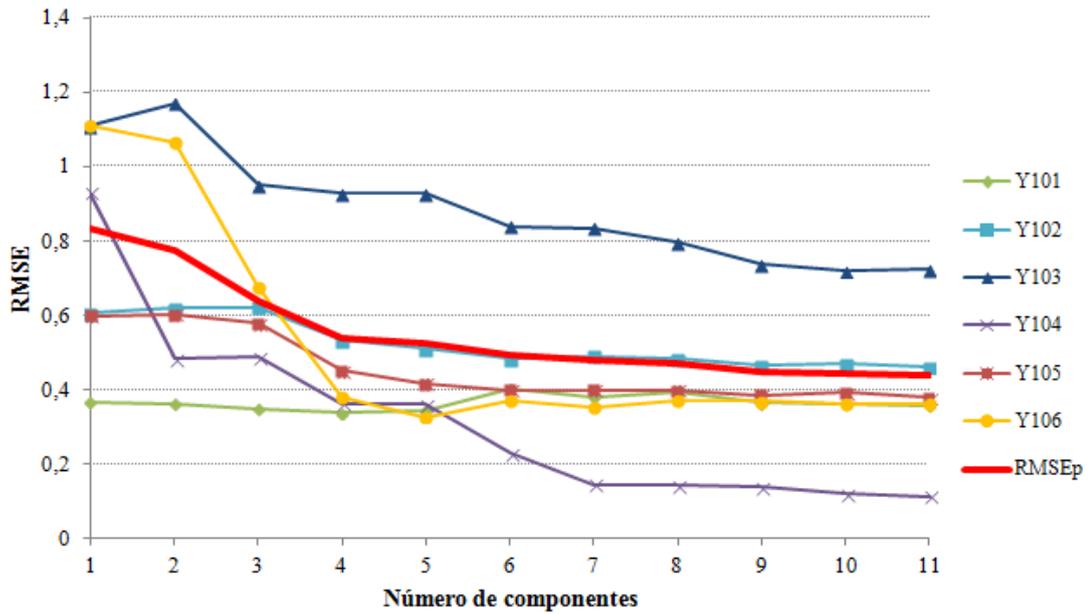
Figura 14 – Determinação do número de componentes para o estágio 1



Fonte: Elaborada pelo autor

A Figura 15 ilustra o comportamento do RMSE para cada variável de resposta do estágio 1 em função do número de componentes para o grupo D. Neste exemplo, observa-se o comportamento não homogêneo no RMSE para cada variável com a adição de componentes retidos na regressão PLS. Algumas variáveis, por exemplo  $y_{106}$ , tem substancial redução no erro quando um maior número de componentes é retido, ao passo que outras, como a  $y_{101}$ , apresentam variação pouco substancial no RMSE. De tal forma, percebe-se que o  $RMSE_p$  é um indicador médio dos erros de predição de cada variável (RMSE), representado o comportamento do sistema com um todo.

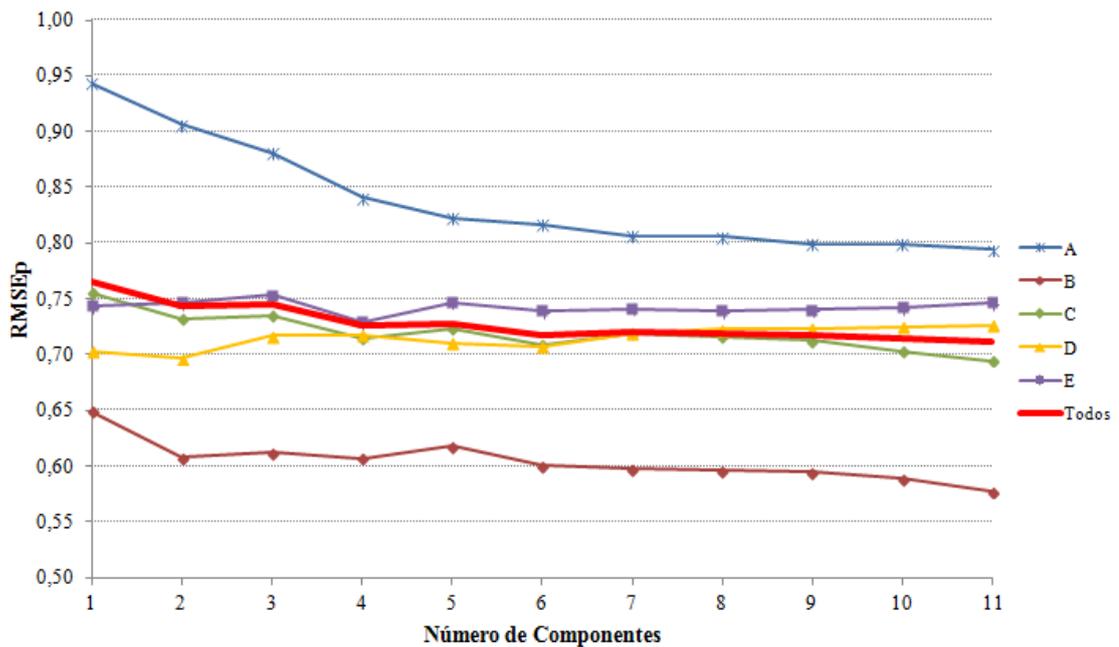
Figura 15 – Comportamento do RMSE em função do número de variáveis latentes para o grupo D do estágio 1



Fonte: Elaborada pelo autor

No estágio 2, o número de componentes escolhido também foi 4, embora dois grupos da validação cruzada indicassem que 2 componentes fossem suficientes, conforme Figura 16.

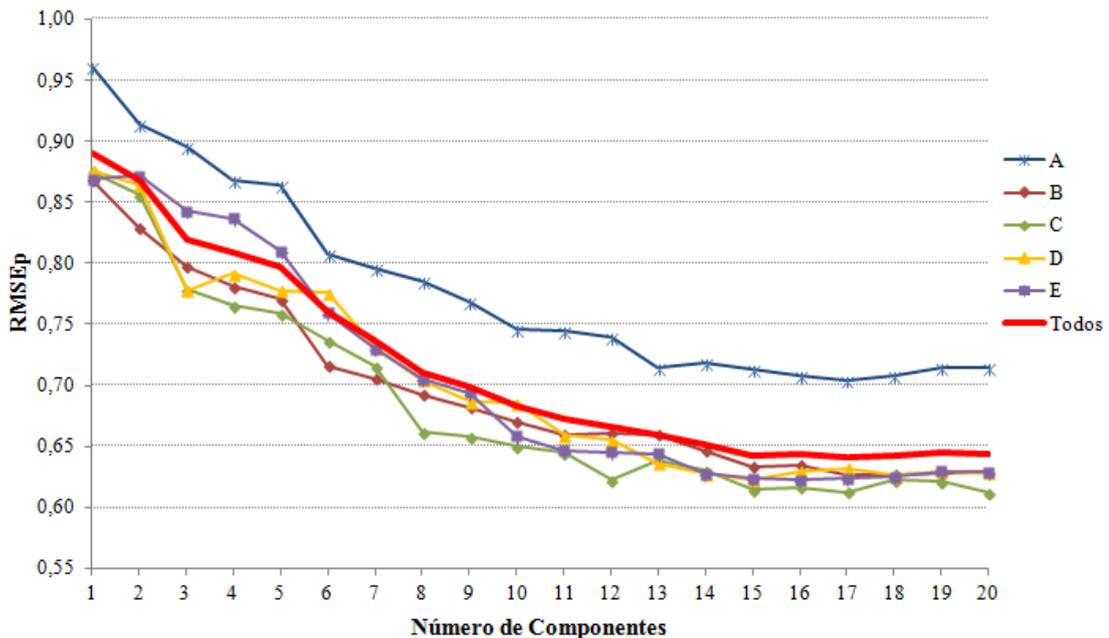
Figura 16 – Determinação do número de componentes para o estágio 2



Fonte: Elaborada pelo autor

No estágio 3 houve divergência no número de componentes, o qual variou de 6 a 10. Observando-se que 2 grupos indicaram 10 componentes, assim como o conjunto agrupado ‘Todos’ (redução de 2,3% no erro de predição ao alterar de 9 para 10 componentes); este foi o número de componentes escolhido, conforme a Figura 17.

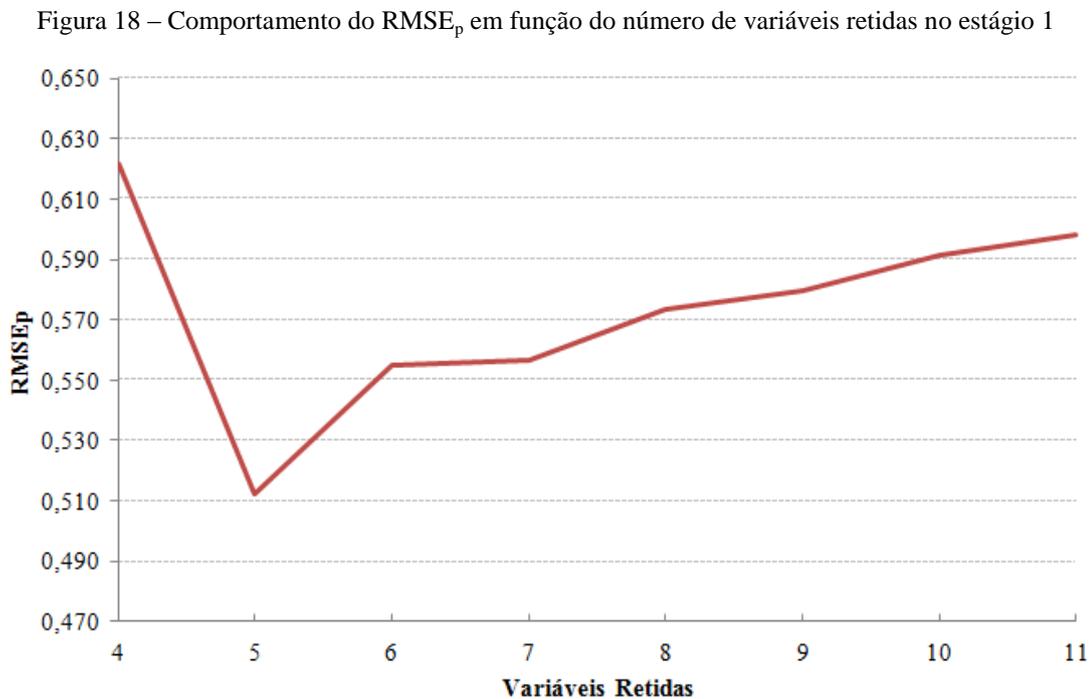
Figura 17 – Determinação do número de componentes para o estágio 3



Fonte: Elaborada pelo autor

Uma vez determinado o número de componentes do modelo PLS para cada estágio, aplicou-se o método proposto para a seleção de variáveis. No primeiro estágio, com 4 componentes, o modelo explica 78,9% da variância de Y ( $R^2_Y=0,789$ ) e reteve 5 das 11 variáveis independentes iniciais:  $x_{101}$ ,  $x_{102}$ ,  $x_{103}$ ,  $x_{104}$  e  $x_{107}$ , as quais descrevem, respectivamente, tempo de processamento, viabilidade, eficiência, custo e volume. Analisando os relacionamentos conhecidos através da opinião de especialista, observa-se que as variáveis tempo de processamento ( $x_{101}$ ) e insumo ( $x_{103}$ ) estão diretamente relacionadas às variáveis de resposta peso médio vivo ( $y_{101}$ ), custo marginal na criação ( $y_{102}$ ) e eficiência ( $y_{105}$ ), viabilidade na criação ( $x_{102}$ ) à viabilidade total ( $y_{103}$ ), custo de processamento ( $x_{104}$ ) a peso médio vivo ( $y_{101}$ ) e custo marginal na criação ( $y_{106}$ ), enquanto que quantidade alojada ( $x_{107}$ ) está relacionado a quantidade abatida ( $y_{104}$ ), o que justifica a retenção destas variáveis no modelo. Entre as três primeiras variáveis eliminadas encontram-se as duas variáveis aleatórias adicionadas ( $x_{110}$  e  $x_{111}$ ), além da variável  $x_{106}$ , que não apresenta correlação

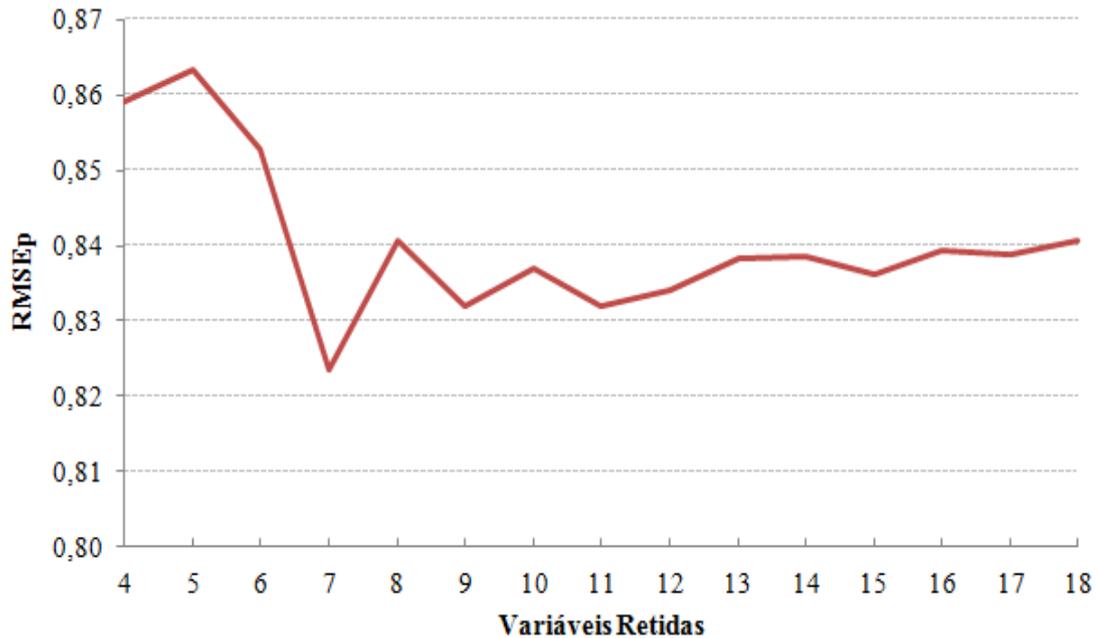
significativa com as variáveis de resposta deste estágio. O comportamento do critério  $RMSE_p$  pode ser observado na Figura 18.



Fonte: Elaborada pelo autor

No segundo estágio, a variância de  $Y$  explicada pelo modelo foi de 48,3% ( $R^2_Y=0,483$ ) e foram retidas 7 das 18 variáveis independentes iniciais. Destas, 3 pertencem ao primeiro estágio ( $x_{101}$ ,  $x_{103}$  e  $x_{104}$ , que já tinham permanecido no primeiro estágio), acrescidas de 4 do segundo estágio ( $x_{201}$ ,  $x_{202}$ ,  $x_{204}$  e  $x_{205}$ ) as quais descrevem, respectivamente, tempo de processamento, temperatura no processo, velocidade da linha e densidade no transporte. A manutenção destas variáveis justifica-se pelo relacionamento existente entre tempo de processamento ( $x_{201}$ ) e temperatura no processo ( $x_{202}$ ) com a variável de resposta temperatura do produto ( $y_{205}$ ), enquanto que velocidade da linha é impactada pelo peso do produto ( $y_{201}$ ) e densidade com peso do produto ( $y_{201}$ ) e peso do subproduto ( $y_{206}$ ). As quatro variáveis aleatórias,  $x_{110}$  e  $x_{111}$  no primeiro estágio e  $x_{206}$  e  $x_{207}$  no segundo estágio figuraram entre as 5 primeiras a serem eliminadas. A Figura 19 apresenta o comportamento do  $RMSE_p$  para este estágio.

Figura 19 – Comportamento do  $RMSE_p$  em função do número de variáveis retidas no estágio 2

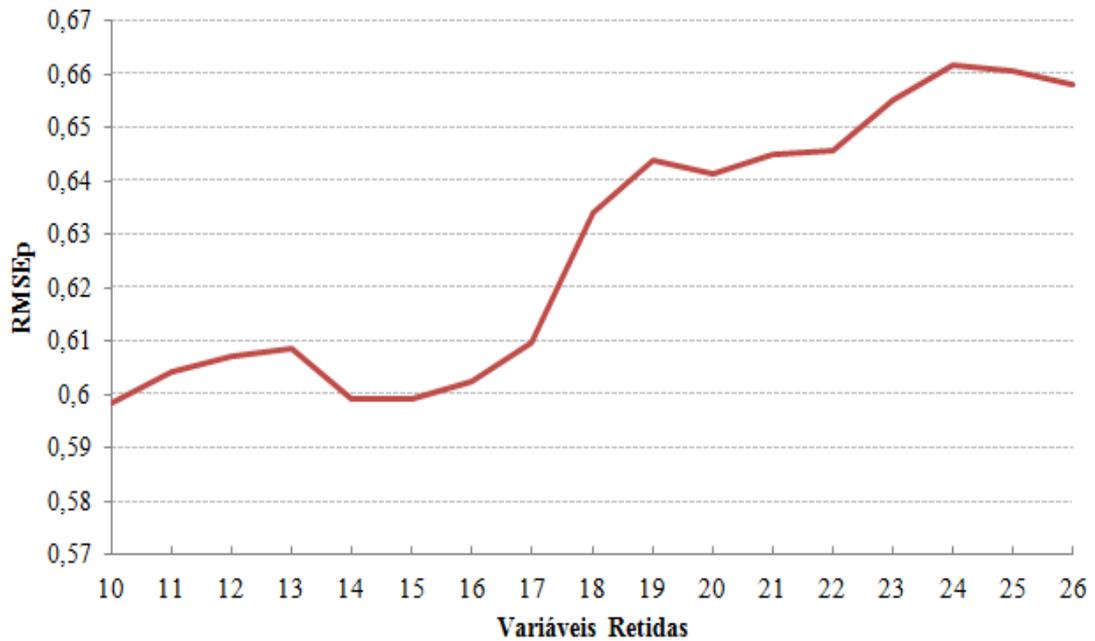


Fonte: Elaborada pelo autor

No terceiro estágio foram retidas 10 das 26 variáveis independentes: 2 do primeiro estágio ( $x_{101}$  e  $x_{103}$ ), 3 do segundo ( $x_{201}$ ,  $x_{202}$  e  $x_{205}$ ) e 5 do último ( $x_{301}$ ,  $x_{302}$ ,  $x_{303}$ ,  $x_{304}$  e  $x_{305}$ ). A variância de Y explicada pelo modelo foi de 68,9% ( $R^2_Y=0,689$ ). Das variáveis do terceiro estágio, peso da perna ( $x_{301}$ ) tem relação com as variáveis de resposta rendimento dos produtos coxa ( $y_{304}$ ) e sobrecoxa ( $y_{305}$ ); peso de peito ( $x_{302}$ ) relaciona-se com rendimentos dos produtos filé de peito ( $y_{301}$ ), peito ( $y_{302}$ ) e retalhos de peito ( $y_{303}$ ); peso de pescoço ( $x_{303}$ ) com rendimento de pescoço ( $y_{308}$ ); peso de gordura ( $x_{304}$ ) com rendimento do produto peito ( $y_{302}$ ) e peso de asa ( $x_{305}$ ) com o rendimento dos produtos meio da asa ( $y_{306}$ ) e ponta da asa ( $y_{307}$ ), o que justifica o fato de ficarem retidas. A Figura 20 mostra a variação em  $RMSE_p$  para o terceiro estágio.

É importante enfatizar que as 6 variáveis aleatórias adicionadas nos três estágios foram eliminadas na primeiras iterações do método proposto (primeiras 10 iterações), o que evidencia a capacidade do modelo em remover variáveis independentes irrelevantes à predição das variáveis de resposta.

Figura 20 – Comportamento do RMSEp em função do número de variáveis retidas no estágio 3



Fonte: Elaborada pelo autor

A Tabela 1 apresenta um resumo dos percentuais de variáveis retidas para cada estágio; em média, foram retidas 40,9% das variáveis originais, o que sugere bons resultados do método proposto em termos de redução do número de variáveis necessárias ao modelo PLS.

Tabela 1 – Percentual de variáveis retidas por estágio

Estágio	Variáveis Independentes Iniciais	Variáveis Retidas	% Variáveis Retidas
1	11	5	45,4
2	18	7	38,9
3	26	10	38,4

Fonte: Elaborada pelo autor

As Tabelas 2 a 4 apresentam os coeficientes de regressão PLS associados às variáveis retidas para cada estágio, com todas as variáveis normalizadas. Note que cada variável de resposta dá origem a um vetor de coeficientes atrelados às variáveis independentes.

Tabela 2 – Coeficientes de regressão PLS do estágio 1

Var. retidas	Variável de resposta					
	y <sub>101</sub>	y <sub>102</sub>	y <sub>103</sub>	y <sub>104</sub>	y <sub>105</sub>	y <sub>106</sub>
x <sub>101</sub>	0,302	0,627	0,241	0,012	0,638	-0,193
x <sub>102</sub>	-0,133	0,287	0,579	-0,092	0,202	0,087
x <sub>103</sub>	0,415	0,382	-0,263	-0,010	0,471	-0,295
x <sub>104</sub>	0,338	-0,299	-0,046	-0,001	-0,408	1,029
x <sub>107</sub>	-0,049	0,059	0,071	0,979	0,046	0,012

Fonte: Elaborada pelo autor

Tabela 3 – Coeficientes de regressão PLS do estágio 2

Var. retidas	Variável de resposta					
	y <sub>201</sub>	y <sub>202</sub>	y <sub>203</sub>	y <sub>204</sub>	y <sub>205</sub>	y <sub>206</sub>
x <sub>101</sub>	-0,008	0,863	0,045	0,132	0,033	-0,129
x <sub>103</sub>	0,382	0,160	-0,220	-0,275	0,095	0,404
x <sub>104</sub>	0,210	-0,279	-0,042	0,124	0,004	0,090
x <sub>201</sub>	-0,037	0,030	-0,068	0,033	-0,557	-0,074
x <sub>202</sub>	-0,086	0,169	0,105	0,030	0,374	-0,066
x <sub>204</sub>	-0,147	0,012	0,006	-0,239	0,135	0,040
x <sub>205</sub>	-0,378	0,049	0,183	0,138	-0,044	-0,336

Fonte: Elaborada pelo autor

Tabela 4 – Coeficientes de regressão PLS do estágio 3

Var. retidas	Variável de resposta									
	y <sub>301</sub>	y <sub>302</sub>	y <sub>303</sub>	y <sub>304</sub>	y <sub>305</sub>	y <sub>306</sub>	y <sub>307</sub>	y <sub>308</sub>	y <sub>309</sub>	y <sub>310</sub>
x <sub>101</sub>	-0,047	0,086	-0,102	0,056	-0,051	-0,089	0,009	-0,012	0,904	-0,058
x <sub>103</sub>	0,039	-0,122	-0,207	-0,211	0,091	0,050	-0,117	-0,026	0,351	0,088
x <sub>201</sub>	-0,018	0,006	0,049	0,072	-0,056	-0,022	0,004	0,012	-0,024	-0,587
x <sub>202</sub>	-0,012	-0,012	0,130	-0,109	0,119	-0,033	0,079	0,003	0,004	0,342
x <sub>205</sub>	-0,176	0,060	0,050	-0,054	0,159	0,221	-0,073	0,021	-0,228	-0,128
x <sub>301</sub>	-0,482	-0,743	0,076	1,530	1,466	-0,680	-0,250	-0,390	-0,093	0,071
x <sub>302</sub>	0,407	1,909	0,184	-1,099	-0,972	-0,698	-0,880	-0,668	-0,386	0,100
x <sub>303</sub>	-0,067	-0,197	-0,155	-0,103	-0,219	-0,127	-0,145	1,258	0,024	-0,174
x <sub>304</sub>	-0,645	-0,664	0,038	0,023	0,007	0,082	-0,158	-0,011	0,088	-0,220
x <sub>305</sub>	0,063	-0,172	0,209	-0,145	-0,121	1,420	0,834	-0,098	-0,403	-0,090

Fonte: Elaborada pelo autor

De tal forma, a predição da variável y<sub>101</sub>, por exemplo, apoia-se na equação (25):

$$\hat{y}_{101} = 0,302 x_{101} - 0,133 x_{102} + 0,415 x_{103} + 0,338 x_{104} - 0,049 x_{107} \quad (25)$$

Selecionadas as variáveis independentes para cada regressão PLS, partiu-se para a definição dos ajustes ótimos destas variáveis. Para tanto, levantaram-se as especificações para as variáveis de resposta bem como do índice de importância dessas variáveis, conforme a Tabela 5. No Apêndice C são apresentados ainda os limites de controle e os índices de capacidade  $C_p$  e  $C_{pk}$ ; pode-se observar que praticamente todas as variáveis apresentam índices menores do que 1, indicando incapacidade de atendimento às especificações. Percebe-se ainda que somente uma variável apresenta índice igual a 1, enquanto que 12 variáveis estão centradas, pois apresentam índices iguais.

Tabela 5 – Limites de especificação das variáveis de resposta e índice de importância

Variável	Limite Inferior de Especificação (LIE)	Valor Nominal ( $\tau$ )	Limite Superior de Especificação (LSE)	Índice Importância ( $I_{qm}$ )
$y_{101}$	15500	16750	18000	1
$y_{102}$	1,40	1,50	1,60	1
$y_{103}$	3,00	12,00	21,00	1
$y_{104}$	5860	7100	8330	1
$y_{105}$	2,50	2,90	3,30	1
$y_{106}$	100	130	160	1
$y_{201}$	13000	14050	15100	1
$y_{202}$	2,00	2,15	2,30	1
$y_{203}$	0,00	7,00	14,00	1
$y_{204}$	0,00	11,00	22,00	1
$y_{205}$	4,00	6,00	8,00	1
$y_{206}$	1000	1300	1600	1
$y_{301}$	4,50	5,65	6,80	1
$y_{302}$	22,50	27,00	31,50	1
$y_{303}$	2,70	3,40	4,10	1
$y_{304}$	13,20	15,50	17,80	1
$y_{305}$	9,50	12,00	14,50	1
$y_{306}$	4,50	5,60	6,70	1
$y_{307}$	4,00	5,00	6,00	1
$y_{308}$	4,00	5,00	6,00	1
$y_{309}$	2,30	2,85	3,40	1
$y_{310}$	4,00	7,00	10,00	1

Fonte: Elaborada pelo autor

Os limites de operação das variáveis independentes foram definidos por especialistas de processo apoiando-se nas seguintes suposições: (i) mínimo observado ou limite inferior de controle, o que for maior; (ii) máximo observado ou limite superior de controle, o que for menor; (iii) utilizou-se a média para a variável  $x_{102}$  (viabilidade na criação), visto que a mesma não é totalmente controlável, mas considerada importante pela empresa; (iv) para a variável  $x_{107}$  (quantidade alojada), a fim de facilitar a interpretação dos resultados, foi escolhido um valor de referência modular; e (v) para a variável  $x_{202}$  (temperatura no processo) foi arbitrado o valor máximo para atendimento da legislação. Esses limites de operação definidos, bem como a média e o desvio-padrão das variáveis independentes retidas, são apresentados na Tabela 6.

Tabela 6 – Limites de operação, média e desvio padrão das variáveis de independentes retidas

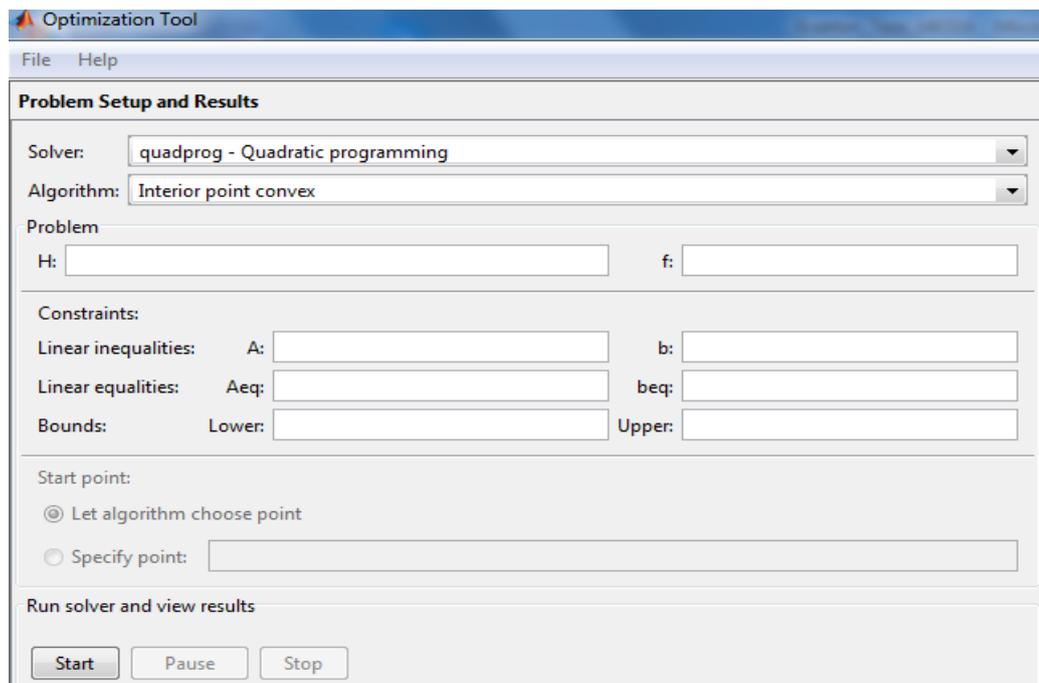
Variável	Limite Inferior de Operação (LIO)	Limite Superior de Operação (LSO)	Média	Desvio-padrão
$x_{101}$	120,00	195,00	162,40	14,58
$x_{102}$	4,40	4,50	4,42	2,77
$x_{103}$	24,50	60,00	42,03	7,71
$x_{104}$	1,10	3,10	2,14	0,39
$x_{107}$	6650,00	9465,60	5991,39	2960,64
$x_{201}$	110,00	130,00	117,16	3,99
$x_{202}$	1,20	5,00	4,35	1,15
$x_{204}$	1,80	2,700	2,16	0,19
$x_{205}$	6,00	12,00	8,68	1,03
$x_{301}$	2950	5250	3842,88	474,22
$x_{302}$	4000	8150	6052,06	803,12
$x_{303}$	490	920	676,45	81,42
$x_{304}$	275	1050	689,31	159,72
$x_{305}$	1220	1790	1477,95	130,70

Fonte: Elaborada pelo autor

De posse dos limites de especificação (Tabela 5), dos limites de operação (Tabela 6) e das regressões PLS para cada estágio (Tabelas 2 a 4), elaborou-se o modelo para otimização. Atribuiu-se a mesma importância para todas as variáveis independentes, ou seja, o termo  $I_{ij}$  na equação (23) é igual a 1 para todos os estágios, conforme a Tabela 5. As regressões foram transformadas da escala normalizada (ver equação 24) para as escalas originais das variáveis.

O modelo final para otimização é apresentado no Apêndice D. A otimização foi realizada através da ferramenta de otimização *Optimtool* do Matlab®, executando a rotina de Programação Quadrática (*quadprog*). A Figura 21 apresenta a tela de entrada da ferramenta *Optimtool*.

Figura 21 – Tela de entrada da ferramenta *Optimization Tool* do Matlab



Fonte: Matlab

A Tabela 7 apresenta os valores previstos para as variáveis de resposta quando os níveis ótimos das variáveis independentes apresentados na Tabela 8 são utilizados. Analisando a Tabela 7, percebe-se que das 22 variáveis de resposta 18 apresentam valores esperados próximos aos nominais, com uma diferença máxima de 3% do valor nominal. Analisando em relação aos limites de especificação, observa-se que o valor previsto mais distante do nominal está a menos de 27% da faixa permitida de variação (diferença entre limite de especificação e valor nominal), enquanto que, para 19 variáveis de resposta esta diferença valor previsto e nominal é menor do que 15% da faixa de variação.

Da tabela 8 observa-se que a condição no processo para a variável  $x_{102}$  (viabilidade na criação) foi a mínima disponibilizada, em função de esta variável ser do tipo ‘menor é melhor’; para a variável  $x_{202}$  (temperatura no processo) o valor encontrado na solução ótima foi o limite máximo disponibilizado; para todas as outras variáveis de processo, os valores previstos situam-se dentro da faixa de operação disponibilizada.

Tabela 7 – Valores esperados e nominais para variáveis de resposta com o modelo proposto

Variável de Resposta	Descrição	Valor esperado	Valor nominal	Unidade
y <sub>101</sub>	Peso (produto)	16621,80	16750,00	G
y <sub>102</sub>	Custo marginal na criação	1,53	1,50	\$/kg
y <sub>103</sub>	Viabilidade total	14,38	12,00	%
y <sub>104</sub>	Quantidade abatida	7077,70	7100,00	g
y <sub>105</sub>	Eficiência	2,85	2,90	g/g
y <sub>106</sub>	Custo (parcela)	134,54	130,00	\$/ton
y <sub>201</sub>	Peso (produto)	14017,00	14050,00	g
y <sub>202</sub>	Custo (carcaça)	2,13	2,15	\$/kg
y <sub>203</sub>	Eficiência (parcial)	6,47	7,00	%
y <sub>204</sub>	Eficiência (processo)	11,55	11,00	%
y <sub>205</sub>	Temperatura carcaça na saída	6,21	6,00	°C
y <sub>206</sub>	Peso (subproduto)	1310,70	1300,00	G
y <sub>301</sub>	Rendimento (PF)	5,63	5,65	%
y <sub>302</sub>	Rendimento (PI)	26,72	27,00	%
y <sub>303</sub>	Rendimento (RP)	3,45	3,40	%
y <sub>304</sub>	Rendimento (S)	15,52	15,50	%
y <sub>305</sub>	Rendimento (C)	11,85	12,00	%
y <sub>306</sub>	Rendimento (AM)	5,59	5,60	%
y <sub>307</sub>	Rendimento (AP)	4,99	5,00	%
y <sub>308</sub>	Rendimento (U)	4,97	5,00	%
y <sub>309</sub>	Custo após processamento	2,88	2,85	\$/kg
y <sub>310</sub>	Temperatura P	6,65	7,00	°C

Fonte: Elaborada pelo autor

Tabela 8 - Condição ótima de trabalho para o processo, gerada pelo modelo proposto

Variável Independente	Descrição	Valor para ajuste	Unidade
x <sub>101</sub>	Tempo de processamento	159,91	dias
x <sub>102</sub>	Viabilidade na criação	4,40	%
x <sub>103</sub>	Insumo (ração/ave)	41,00	kg
x <sub>104</sub>	Custo no processamento	2,25	\$/kg
x <sub>107</sub>	Quantidade alojada	8329,87	ave
x <sub>201</sub>	Tempo de processamento	123,47	min
x <sub>202</sub>	Temperatura (processo)	5,00	°C
x <sub>204</sub>	Velocidade da linha	2,11	mil ave/h
x <sub>205</sub>	Densidade no transporte	8,26	ave/cx
x <sub>301</sub>	Peso perna (L)	3842,14	g
x <sub>302</sub>	Peso peito (P)	6061,44	g
x <sub>303</sub>	Peso pescoço (U)	695,53	g
x <sub>304</sub>	Peso gordura (G)	690,62	g
x <sub>305</sub>	Peso asa(A)	1480,56	g

Fonte: Elaborada pelo autor

A análise pelo especialista da empresa das Tabelas 7 e 8 mostra que os resultados são coerentes.

Para fins de comparação com os resultados obtidos através da sistemática de seleção de variáveis proposta, gerou-se um modelo de regressão PLS no qual todas as variáveis independentes são mantidas, com o mesmo número de componentes definidos em cada estágio (4, 4 e 10). De posse do modelo, foi elaborada uma formulação de otimização utilizando os mesmos limites de especificação e critérios de definição dos limites de operação, aplicados anteriormente. Este modelo será chamado de ‘modelo completo’.

A Tabela 9 apresenta indicadores dos resultados entre os modelos completo e o proposto, onde se observa que embora o valor da Função Objetivo do modelo completo seja inferior ao modelo proposto, a capacidade de predição do modelo proposto é melhor em todos os estágios (assim como o coeficiente de determinação é maior no modelo proposto).

Tabela 9 – Comparativo de resultados entre os modelos completo e proposto

Indicador	Estágio	Modelo Completo	Modelo Proposto	Diferença
RMSEp	1	0,598	0,512	0,086
RMSEp	2	0,841	0,824	0,017
RMSEp	3	0,658	0,598	0,600
R <sup>2</sup> Y	1	73,9%	78,9%	5,0%
R <sup>2</sup> Y	2	47,2%	48,3%	1,1%
R <sup>2</sup> Y	3	64,1%	68,9%	4,8%
F. Objetivo		0,0435	0,0634	0,0199

Fonte: Elaborada pelo autor

Na Tabela 10 são apresentados os resultados desta simulação para as variáveis de resposta, bem como os resultados obtidos através da seleção (apresentados na Tabela 7). Observa-se que as diferenças entre os valores previstos pelos dois modelos são marginais, na grande maioria, com exceção das variáveis  $y_{103}$  (viabilidade total) cuja diferença é de cerca de 20% e  $y_{203}$  (eficiência parcial) com 9% de diferença.

Na Tabela 11 estão os resultados para as variáveis independentes obtidas pelo modelo completo e os obtidos através da sistemática de seleção proposta. A análise de especialista identificou algumas incoerências no resultado do modelo completo: por exemplo, o valor esperado para a variável de resposta  $y_{101}$  (peso do produto) é baixo, considerando o tempo de processamento de 170,37 (variável  $x_{101}$ ), e os valores elevados para as variáveis  $x_{107}$

(quantidade alojada) e  $x_{102}$  (viabilidade na criação) e a respectiva variável de resposta  $y_{104}$  (quantidade abatida), que se manteve igual.

Tabela 10 – Comparativo dos valores esperados para as variáveis de resposta pelo modelo completo e proposto

Variável de Resposta	Descrição	Completo	Proposto	Valor nominal	Diferença
$y_{101}$	Peso (produto)	16750,00	16621,80	16750,00	-128,20
$y_{102}$	Custo marginal na criação	1,52	1,53	1,50	0,03
$y_{103}$	Viabilidade total	11,99	14,38	12,00	2,38
$y_{104}$	Quantidade abatida	7100,00	7077,70	7100,00	-22,30
$y_{105}$	Eficiência	2,86	2,85	2,90	-0,01
$y_{106}$	Custo (parcela)	129,99	134,54	130,00	4,54
$y_{201}$	Peso (produto)	14049,59	14017,00	14050,00	-32,46
$y_{202}$	Custo (carcaça)	2,23	2,13	2,15	-0,10
$y_{203}$	Eficiência (parcial)	7,13	6,47	7,00	-0,66
$y_{204}$	Eficiência (processo)	11,26	11,55	11,00	0,29
$y_{205}$	Temperatura.carcaça na saída	5,91	6,21	6,00	0,30
$y_{206}$	Peso (subproduto)	1301,92	1310,70	1300,00	8,79
$y_{301}$	Rendimento (PF)	5,60	5,63	5,65	0,03
$y_{302}$	Rendimento (PI)	27,01	26,72	27,00	-0,29
$y_{303}$	Rendimento (RP)	3,39	3,45	3,40	0,05
$y_{304}$	Rendimento (S)	15,53	15,52	15,50	-0,01
$y_{305}$	Rendimento (C)	11,94	11,85	12,00	-0,09
$y_{306}$	Rendimento (AM)	5,63	5,59	5,60	-0,04
$y_{307}$	Rendimento (AP)	5,02	4,99	5,00	-0,03
$y_{308}$	Rendimento (U)	5,00	4,97	5,00	-0,03
$y_{309}$	Custo após processamento	2,85	2,88	2,85	0,03
$y_{310}$	Temperatura P	7,02	6,65	7,00	-0,37

Fonte: Elaborada pelo autor

Outro ponto que chama a atenção na Tabela 11 são os valores das variáveis artificiais (que apresentam média próxima a 0 e desvio padrão ao redor de 1), mas na condição ótima no modelo completo os valores sugeridos são de +2,58 para a variável  $x_{111}$  (artificial 2), que é o limite superior de operação; 2,62 para a variável  $x_{207}$  (artificial 4), que é o limite inferior de operação e -2,17 e -2,40, para as variáveis  $x_{307}$  (artificial 5) e  $x_{308}$  (artificial 6), valores bem próximos aos limites inferiores destas variáveis.

Tabela 11 - Condição ótima de trabalho para o processo nos modelos completo e proposto

Variável Independente	Descrição	Completo	Proposto	Diferença
$x_{101}$	Tempo de processamento	170,37	159,91	10,46
$x_{102}$	Viabilidade na criação	4,40	4,40	0,00
$x_{103}$	Insumo (ração/ave)	45,36	41,00	4,36
$x_{104}$	Custo no processamento	1,56	2,25	0,69
$x_{105}$	Despesa no processamento	1,19		
$x_{106}$	Tempo de espera	1,66		
$x_{107}$	Quantidade alojada	9450,23	8329,87	1120,37
$x_{108}$	Transporte	14,77		
$x_{109}$	Densidade alojada	2,58		
$x_{110}$	Artificial 1	-0,62		
$x_{111}$	Artificial 2	2,58		
$x_{201}$	Tempo de processamento	117,89	123,47	5,57
$x_{202}$	Temperatura (processo)	4,99	5,00	0,01
$x_{203}$	Temperatura (produto)	15,64		
$x_{204}$	Velocidade da linha	2,23	2,21	0,02
$x_{205}$	Densidade no transporte	7,86	8,26	0,40
$x_{206}$	Artificial 3	1,51		
$x_{207}$	Artificial 4	-2,62		
$x_{301}$	Peso perna (L)	4321,11	3842,14	478,98
$x_{302}$	Peso peito (P)	7355,58	6061,44	1294,13
$x_{303}$	Peso pescoço (U)	651,28	695,53	44,25
$x_{304}$	Peso gordura (G)	278,10	690,62	412,51
$x_{305}$	Peso asa(A)	1265,56	1480,56	215,00
$x_{306}$	Velocidade espotejamento	29,04		
$x_{307}$	Artificial 5	-2,17		
$x_{308}$	Artificial 6	-2,40		

Fonte: Elaborada pelo autor

## 4.2 EXPERIMENTO SIMULADO

O banco de dados utilizado no experimento simulado foi baseado em observações de um processo industrial de polimerização para fabricação de látex. A base original apresenta 262 observações, com 117 variáveis independentes, contendo medidas de temperaturas, tempos, taxa de entrada do monômero, nível de catalisador, concentração reativa, entre outras; o banco apresenta somente uma variável de resposta, que mede a quantidade de produtos secundários insolúveis. Esta base foi utilizada por Gauchi e Chagnon (2001) para comparar

métodos de seleção de variáveis com objetivo de predição de variáveis, enquanto Anzanello *et al.* (2009) a utilizaram para comparar métodos de seleção de variáveis com o objetivo de classificação de bateladas produtivas.

Gauchi e Chagnon (2001) enfatizam a dificuldade de gerar um modelo preditivo em dados com variáveis independentes altamente correlacionadas. Em função desta característica, esta base foi escolhida para este estudo simulado, de forma a avaliar o comportamento do método proposto. Tendo em vista que o processo original ocorre em único estágio, as variáveis independentes foram divididas em 3 partes iguais (com 39 variáveis), simulando três estágios. Foram geradas 8 variáveis de resposta para cada estágio, com diversos graus de relacionamento com algumas variáveis independentes, conforme descrito na Tabela 12. Estas variáveis aleatórias (num total de 23, com 262 observações) foram geradas através da opção *Multivariate Normal Distribution* no software Minitab 14.1. Cada variável de resposta simulada foi relacionada a uma ou mais variáveis independentes, do banco de dados original (Latex), adicionando uma variável aleatória gerada, ponderada de forma a gerar um coeficiente de correlação desejado pré-definido em quatro níveis: 0,30 (correlação pequena); 0,50 (correlação moderada); 0,75 (correlação forte) e 0,90 (correlação muito forte), porém, estes coeficientes, apresentados na Tabela 12, alteraram um pouco devido à etapa de pré-tratamento descrita adiante nesta seção.

Ainda de acordo com a Tabela 12, buscou-se, para a variável de resposta  $y_{101}$  do primeiro estágio, uma correlação forte com uma variável independente deste estágio ( $x_{121}$ );  $y_{102}$  foi correlacionada com intensidade muito forte a  $x_{122}$ , enquanto que para  $y_{103}$  uma correlação moderada com  $x_{125}$  foi estabelecida; a variável  $y_{104}$  não foi correlacionada a nenhuma variável independente, sendo ela uma das variáveis aleatórias geradas, com o objetivo de verificar como o modelo se comporta numa situação de correlação inexistente ou leve. Já  $y_{105}$  foi relacionada com igual intensidade a duas variáveis independentes  $x_{106}$  e  $x_{135}$ ;  $y_{106}$  foi relacionada a duas variáveis independentes  $x_{122}$  e  $x_{134}$ , mas com correlações pequenas;  $y_{107}$  foi associada a quatro variáveis independentes, com correlações moderadas ( $x_{102}$ ;  $x_{106}$  e  $x_{136}$ ) e pequena com  $x_{134}$ ;  $y_{108}$  foi associada a três variáveis independentes, as mesmas de correlações moderadas usadas em  $y_{107}$ .

Para o segundo estágio, tem-se a variável de resposta  $y_{201}$  fortemente relacionada a uma variável independente do segundo estágio;  $y_{202}$  com uma correlação moderada com uma variável do segundo estágio;  $y_{203}$  relacionada de forma moderada com uma variável do primeiro estágio e fortemente com uma variável do segundo estágio;  $y_{204}$  relacionada a duas

variáveis do primeiro estágio (correlação moderada e forte);  $y_{205}$  relacionada a duas variáveis do segundo estágio (correlação moderada e forte);  $y_{206}$  relacionada a três variáveis do segundo estágio (correlações moderada e forte);  $y_{207}$  relacionada a duas variáveis do primeiro estágio (correlação moderada) e  $y_{208}$  relacionada fortemente com uma variável do primeiro estágio e de forma moderada com uma variável do segundo estágio.

Tabela 12– Relacionamentos gerados entre variáveis de resposta e independentes

Variável de Resposta	Variáveis independentes relacionadas	Correlação ( $r$ )
$y_{101}$	$x_{121}$	0,74
$y_{102}$	$x_{122}$	0,88
$y_{103}$	$x_{125}$	0,53
$y_{104}$	nenhuma	-
$y_{105}$	$x_{106}$ ; $x_{135}$	0,63 ; 0,63
$y_{106}$	$x_{122}$ ; $x_{134}$	-0,40 ; 0,37
$y_{107}$	$x_{102}$ ; $x_{106}$ ; $x_{136}$ ; $x_{138}$	0,59 ; 0,50 ; 0,58 ; 0,33
$y_{108}$	$x_{102}$ ; $x_{106}$ ; $x_{136}$	-0,56 ; -0,50 ; -0,73
$y_{201}$	$x_{224}$	-0,72
$y_{202}$	$x_{205}$	-0,48
$y_{203}$	$x_{131}$ ; $x_{208}$	0,42 ; 0,87
$y_{204}$	$x_{114}$ ; $x_{126}$	0,62 ; 0,79
$y_{205}$	$x_{203}$ ; $x_{220}$	0,70 ; -0,73
$y_{206}$	$x_{204}$ ; $x_{208}$ ; $x_{235}$	0,58 ; 0,73 ; 0,42
$y_{207}$	$x_{106}$ ; $x_{135}$	-0,63 ; -0,63
$y_{208}$	$x_{122}$ ; $x_{232}$	0,83 ; 0,61
$y_{301}$	original	-
$y_{302}$	$x_{224}$ ; $x_{324}$	0,78 ; 0,79
$y_{303}$	$x_{326}$ ; $x_{335}$	0,79 ; 0,83
$y_{304}$	$x_{315}$ ; $x_{320}$ ; $x_{321}$	-0,87 ; 0,59 ; 0,29
$y_{305}$	$x_{114}$ ; $x_{126}$ ; $x_{304}$	0,40 ; 0,59 ; 0,66
$y_{306}$	$x_{205}$ ; $x_{310}$	0,43 ; 0,88
$y_{307}$	$x_{122}$ ; $x_{315}$	-0,32 ; 0,89
$y_{308}$	$x_{102}$ ; $x_{220}$ ; $x_{320}$	0,36 ; 0,62 ; 0,90

Fonte: Elaborada pelo autor

No terceiro estágio, para a variável  $y_{301}$  foi mantida a variável de resposta original do látex;  $y_{302}$  relacionada fortemente a uma variável do segundo estágio e outra do terceiro;  $y_{303}$  relacionada a duas variáveis do terceiro estágio (correlação forte);  $y_{304}$  relacionada a três variáveis do terceiro estágio (correlações forte, moderada e pequena);  $y_{305}$  relacionada a três variáveis, sendo duas do primeiro estágio (correlações pequena e moderada) e outra do

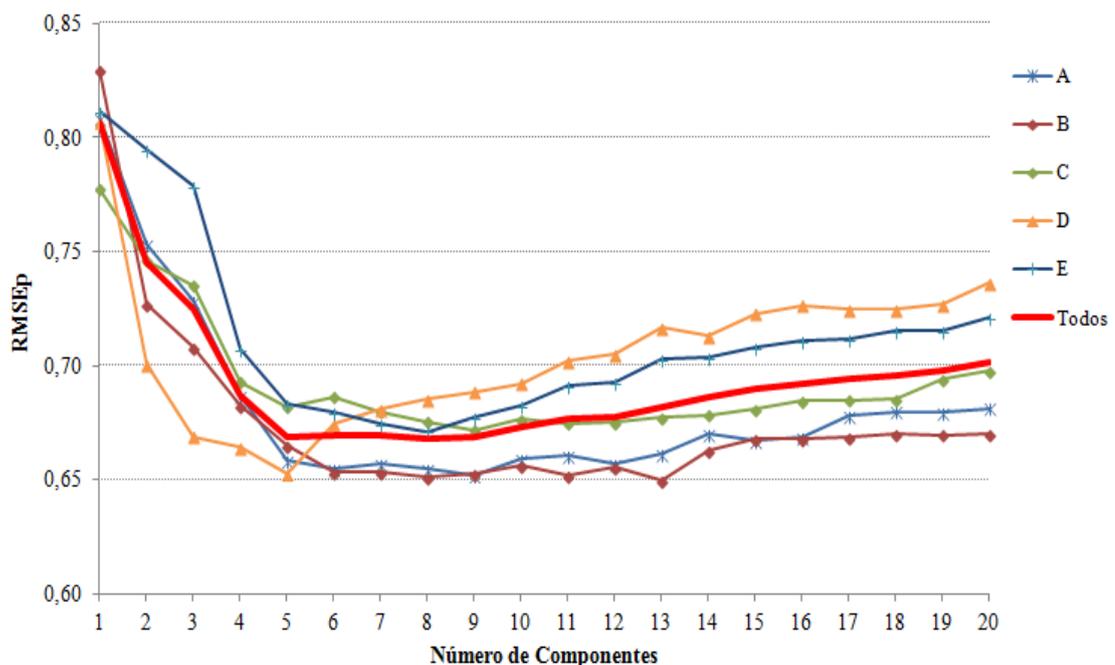
terceiro estágio (correlação moderada);  $y_{306}$  relacionada com intensidade moderada a uma variável do segundo estágio e outra do terceiro estágio (correlação forte);  $y_{307}$  relacionada com intensidade pequena a uma variável do primeiro estágio e outra do terceiro estágio (correlação forte);  $y_{308}$  relacionada a três variáveis, sendo uma do primeiro estágio (correlação pequena), uma do segundo estágio (correlação moderada) e outra do terceiro estágio (correlação forte).

Na sequência, com estágios e variáveis definidas, procedeu-se à etapa de pré-tratamento dos dados, com análise de valores espúrios. Foram retidas 225 das 262 observações originais.

Geraram-se então as regressões PLS para identificar o número de variáveis latentes por estágio. A regressão PLS para o primeiro estágio contou com 39 variáveis independentes ( $x_{101}$  a  $x_{139}$ ) e 8 variáveis de resposta ( $y_{101}$  a  $y_{108}$ ); o segundo estágio partiu de 78 variáveis independentes ( $x_{101}$  a  $x_{139}$ ;  $x_{201}$  a  $x_{239}$ ) e 8 variáveis de resposta ( $y_{201}$  a  $y_{208}$ ); e o terceiro estágio de 117 variáveis independentes ( $x_{101}$  a  $x_{139}$ ;  $x_{201}$  a  $x_{239}$ ;  $x_{301}$  a  $x_{339}$ ) e 8 variáveis de resposta ( $y_{301}$  a  $y_{308}$ ). Utilizou-se o procedimento de validação cruzada com proporção de treino/teste de 80% e 20% e cinco grupos (similar ao apresentado no estudo de caso anterior).

Para o primeiro estágio, o número de componentes escolhido foi 5, conforme análise da Figura 22.

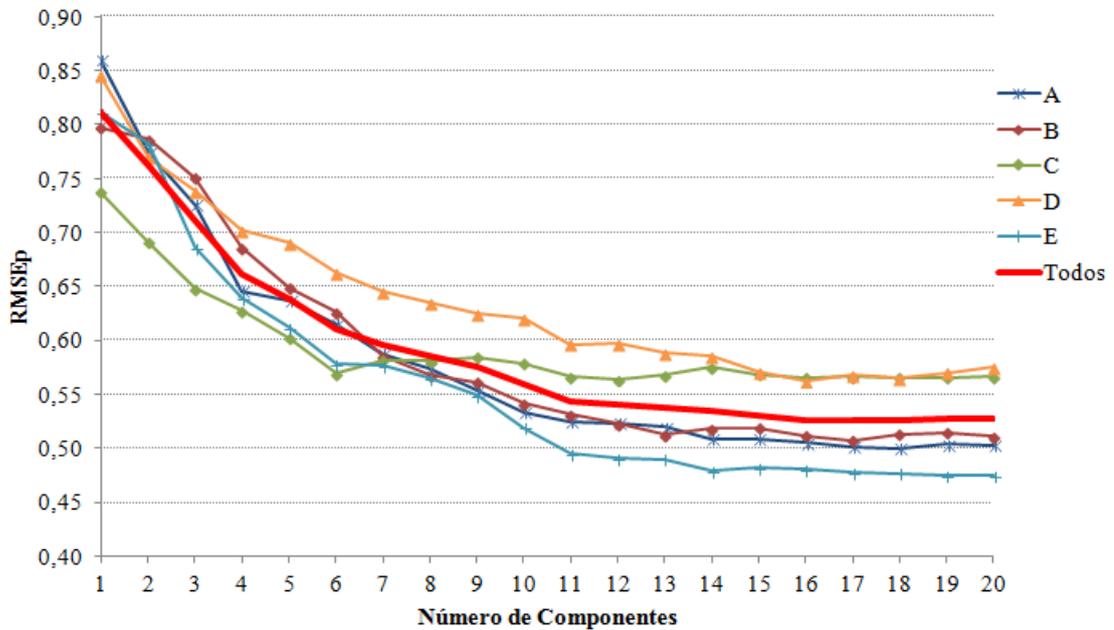
Figura 22 – Determinação do número de componentes para o estágio 1 para o estudo simulado



Fonte: Elaborada pelo autor

Já no segundo estágio, o número de componentes escolhidos foi 11, o que pode ser visto na Figura 23.

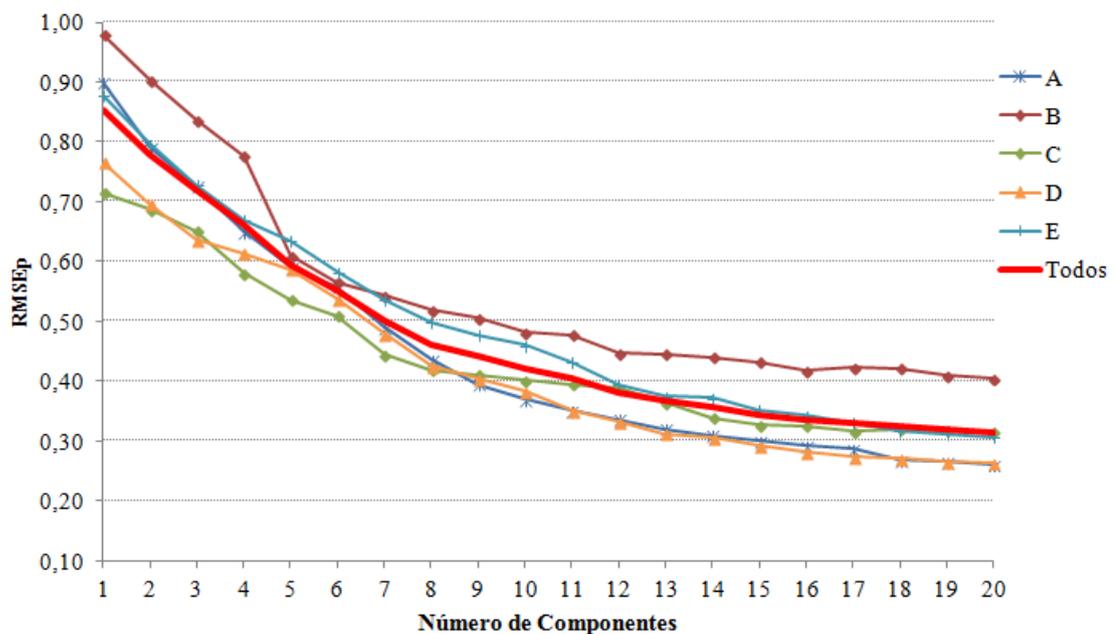
Figura 23 – Determinação do número de componentes para o estágio 2 para o estudo simulado



Fonte: Elaborada pelo autor

No último estágio, foram retidos 12 componentes, mediante análise da Figura 24.

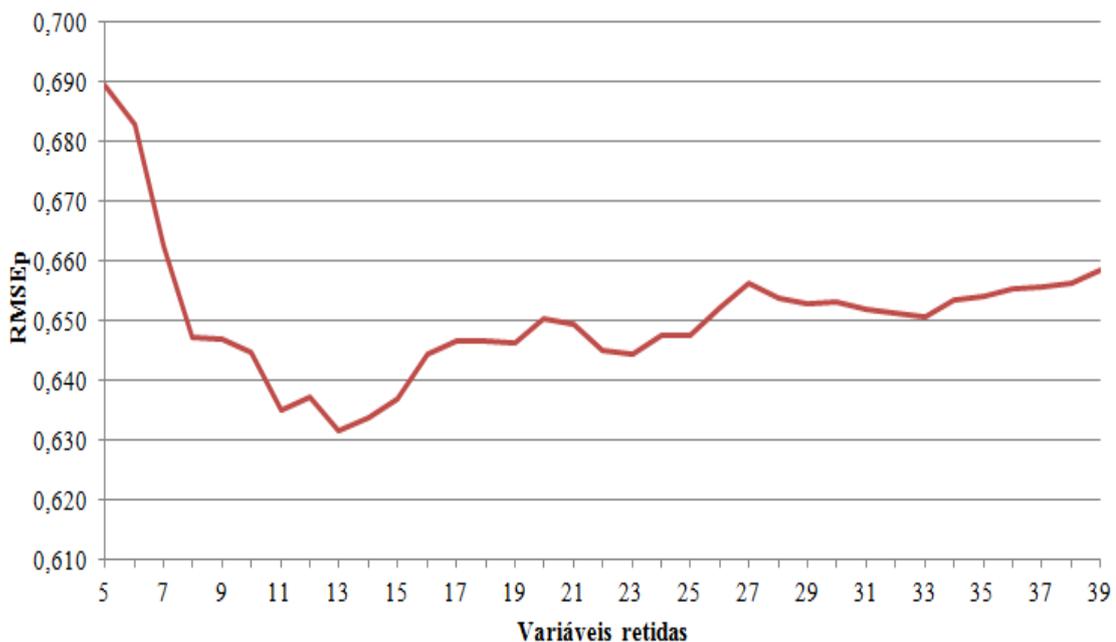
Figura 24 – Determinação do número de componentes para o estágio 3 para o estudo simulado



Fonte: Elaborada pelo autor

Concluída a etapa de seleção de variáveis do método proposto, a variância explicada de Y pelo modelo foi de 60,4% ( $R^2_Y=0,604$ ), o erro médio de previsão  $RMSE_p$  foi de 0,632 e 13 variáveis independentes foram retidas no primeiro estágio, conforme a Figura 25:  $x_{102}$ ,  $x_{106}$ ,  $x_{118}$ ,  $x_{119}$ ,  $x_{121}$ ,  $x_{122}$ ,  $x_{123}$ ,  $x_{125}$ ,  $x_{134}$ ,  $x_{135}$ ,  $x_{136}$ ,  $x_{137}$  e  $x_{138}$ . Observando-se na Tabela 12, nove dessas variáveis foram utilizadas para calcular o valor das variáveis de resposta e, portanto, era esperado que fossem retidas. Contudo, as variáveis  $x_{118}$ ,  $x_{119}$ ,  $x_{123}$  e  $x_{137}$  não foram utilizadas na geração das variáveis de resposta, logo não se esperava que fossem selecionadas; observando-se a matriz de correlação das variáveis (ver Apêndice E), verifica-se que as variáveis  $x_{118}$  e  $x_{119}$  apresentam correlação muito forte ( $r=0,93$ ) e ainda apresentam correlações de moderadas a fortes ( $0,65 \leq r \leq 0,88$ ) com as variáveis  $x_{121}$  e  $x_{122}$ ; também a variável  $x_{123}$  tem correlação muito forte ( $r = 0,93$ ) com estas variáveis ( $x_{121}$  e  $x_{122}$ ); já a variável  $x_{137}$  é apresenta correlação forte com  $x_{136}$  ( $r = 0,73$ ).

Figura 25 –  $RMSE_p$  em função do número de variáveis retidas no estágio 1 para o estudo simulado



Fonte: Elaborada pelo autor

Ao avaliar-se a correlação das variáveis retidas, percebe-se que o conjunto  $x_{118}$ ,  $x_{119}$ ,  $x_{120}$ ,  $x_{121}$ ,  $x_{122}$ ,  $x_{123}$  apresenta correlações muito fortes, fortes e moderadas, conforme pode ser observado na Tabela 13. Três destas variáveis permaneceram no modelo,  $x_{118}$ ,  $x_{119}$  e  $x_{123}$  e se verifica que a correlação entre  $x_{122}$  e  $x_{123}$  é praticamente perfeita ( $r = 1,00$ ), o que caracteriza redundância.

Tabela 13 - Matriz dos coeficientes de correlação de algumas variáveis do estágio 1

	$x_{118}$	$x_{119}$	$x_{120}$	$x_{121}$	$x_{122}$	$x_{123}$
$x_{118}$	1	0,93	0,74	0,76	0,65	0,67
$x_{119}$		1	0,84	0,88	0,75	0,76
$x_{120}$			1	0,99	0,95	0,95
$x_{121}$				1	0,95	0,95
$x_{122}$					1	1,00
$x_{123}$						1

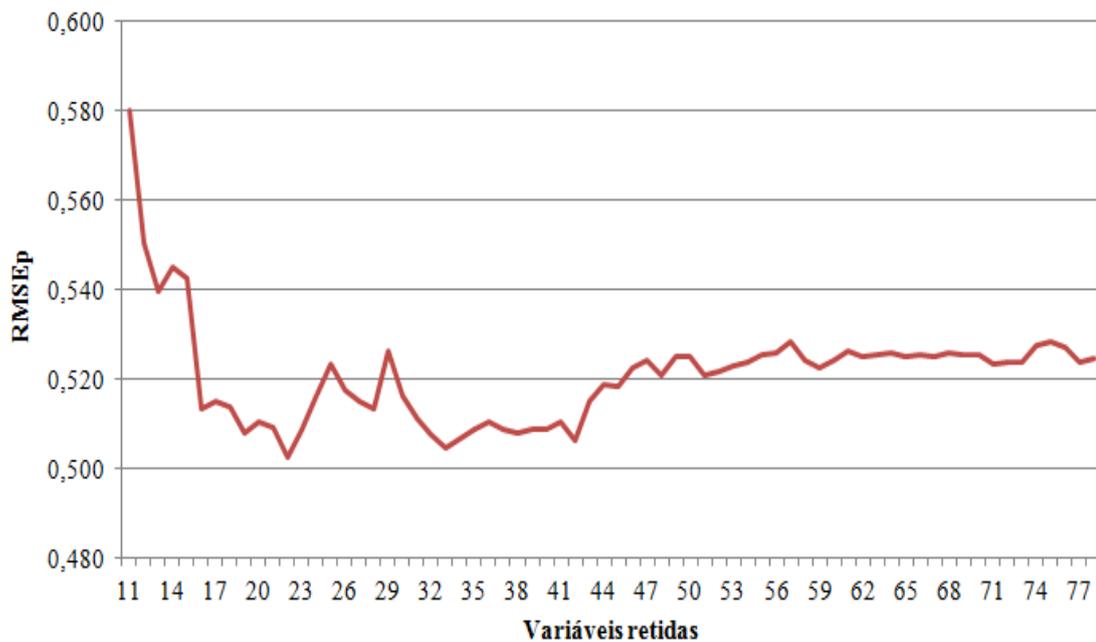
Fonte: Extraída do Apêndice F

Diante deste fato, pode-se questionar o impacto que duas variáveis de correlação muito forte trazem ao modelo. Para avaliar especificamente este impacto, gerou-se um modelo PLS excluindo a variável  $x_{123}$  do conjunto das variáveis independentes; obteve-se um modelo semelhante em termos de erro de previsão ( $RMSE_p = 0,633$ ) e variância explicada de Y de 60,2%. Comparando os coeficientes de regressão PLS (do Apêndice F – Estágio 1) com o novo modelo PLS, observa-se que, para as variáveis de resposta  $y_{101}$ ,  $y_{102}$  e  $y_{106}$ , a soma dos coeficientes de regressão do conjunto de variáveis  $x_{118}$  a  $x_{123}$  foi, respectivamente, de 0,713, 0,773 e -0,572; enquanto que, no modelo onde  $x_{123}$  foi excluída, a soma dos coeficientes de regressão para o conjunto de variáveis retidas ( $x_{119}$ ,  $x_{120}$ ,  $x_{121}$ ,  $x_{122}$  e  $x_{125}$ ) foi 0,715, 0,776 e -0,611 (com alterações de 0%, 0% e 7% na soma, respectivamente). Enfatiza-se ainda que, excluindo a variável  $x_{123}$ , o modelo com menor erro de previsão teria retido 10 variáveis e não 12, com as variáveis  $x_{120}$  e  $x_{137}$  não permanecendo no modelo. Nessas circunstâncias, a soma dos coeficientes de regressão para o conjunto de variáveis analisadas ( $x_{119}$ ,  $x_{121}$ ,  $x_{122}$  e  $x_{125}$ ) para as mesmas variáveis de resposta é 0,705, 0,718 e -0,552, com alterações de 1%, 7% e 3%. De tal forma, sugere-se que o impacto do descarte das variáveis  $x_{118}$  e  $x_{123}$  foi compensado (em grande parte) no modelo pela alteração dos coeficientes de regressão nas variáveis correlacionadas que permaneceram retidas no modelo.

Para o segundo estágio, a Figura 26 apresenta o perfil de erro com o número de variáveis retidas. A variância explicada de Y pelo modelo foi de 77,9% ( $R^2_Y = 0,779$ ) e foram retidas 22 variáveis independentes, sendo 10 do primeiro estágio ( $x_{106}$ ,  $x_{108}$ ,  $x_{109}$ ,  $x_{113}$ ,  $x_{121}$ ,  $x_{125}$ ,  $x_{126}$ ,  $x_{127}$ ,  $x_{131}$  e  $x_{134}$ ) e 12 do segundo estágio ( $x_{203}$ ,  $x_{204}$ ,  $x_{205}$ ,  $x_{207}$ ,  $x_{208}$ ,  $x_{209}$ ,  $x_{212}$ ,  $x_{219}$ ,  $x_{220}$ ,  $x_{222}$ ,  $x_{223}$  e  $x_{224}$ ). Novamente, conforme a Tabela 12 se esperava que fossem retidas as seguintes variáveis do primeiro estágio:  $x_{106}$ ,  $x_{114}$ ,  $x_{122}$ ,  $x_{126}$ ,  $x_{131}$  e  $x_{135}$ , já que as variáveis de resposta foram geradas a partir destas variáveis independentes com correlação moderada ou mais forte

(e apenas três delas foram retidas:  $x_{106}$ ,  $x_{126}$  e  $x_{131}$ ). Em função das correlações muito fortes entre as variáveis  $x_{121}$  e  $x_{122}$  ( $r = 0,95$ ) e forte entre as variáveis  $x_{134}$  e  $x_{135}$  ( $r = 0,86$ ), presume-se que a variável  $x_{122}$  foi substituída no modelo por  $x_{121}$ , bem como  $x_{135}$  substituída por  $x_{134}$ . A variável  $x_{108}$  apresenta correlação moderada ( $r = 0,64$ ) com  $x_{114}$ , enquanto que a variável  $x_{113}$  tem correlação forte ( $r = 0,77$ ) com esta mesma variável (não retida); desta forma,  $x_{114}$  foi substituída pelas variáveis  $x_{108}$  e  $x_{113}$ . O grupo composto pelas variáveis  $x_{109}$ ,  $x_{125}$  e  $x_{127}$  foi retido por redundância, pois se observa correlações moderadas entre  $x_{109}$  e  $x_{220}$  ( $r = -0,68$ ) e entre  $x_{109}$  e  $x_{208}$  ( $r=0,50$ ); a presença de uma correlação muito forte ( $r = 0,97$ ) entre  $x_{125}$  e  $x_{126}$  ( $r=0,97$ ), que ficou retida; e as correlações da variável  $x_{127}$ , de intensidade forte ( $r = -0,78$ ) com  $x_{208}$  e moderada ( $r = 0,69$ ) com  $x_{220}$ .

Figura 26 – RMSEP em função do número de variáveis retidas no estágio 2 para o estudo simulado



Fonte: Elaborada pelo autor

Em relação às variáveis do segundo estágio, a expectativa era que  $x_{203}$ ,  $x_{204}$ ,  $x_{205}$ ,  $x_{208}$ ,  $x_{220}$ ,  $x_{224}$ ,  $x_{232}$  e  $x_{235}$  fossem retidas, pois foram utilizadas para gerar as variáveis de resposta (ver Tabela 12). Destas, 6 foram retidas e 2 não ( $x_{232}$  e  $x_{235}$ ), enquanto que outras 6 não previamente esperadas foram retidas. As variáveis  $x_{207}$  e  $x_{209}$  ficaram retidas por conta de correlação muito forte ( $r = 0,99$ ) com  $x_{208}$ . As variáveis  $x_{212}$  e  $x_{219}$  apresentam correlação moderada ( $r = 0,50$  e  $0,44$ , respectivamente) com  $x_{235}$ , possivelmente substituindo-a por tal razão. A variável  $x_{222}$  tem correlação moderada ( $r = 0,53$ ) com  $x_{135}$ , enquanto que uma

correlação pequena ( $r = -0,40$ ) é observada de  $x_{223}$  com  $x_{205}$  ( $r = -0,40$ ). Para a variável de resposta  $y_{208}$ , observa-se que houve a substituição de  $x_{122}$  por  $x_{121}$  (correlação muito forte,  $r = 0,95$ ) e, por fim, a variável  $x_{232}$ , mesmo com uma correlação moderada ( $r = 0,61$ ), não foi retida pela falta de correlação ( $r \leq 0,20$ ) com as outras variáveis de resposta, sendo assim substituída por  $x_{121}$  e  $x_{125}$ .

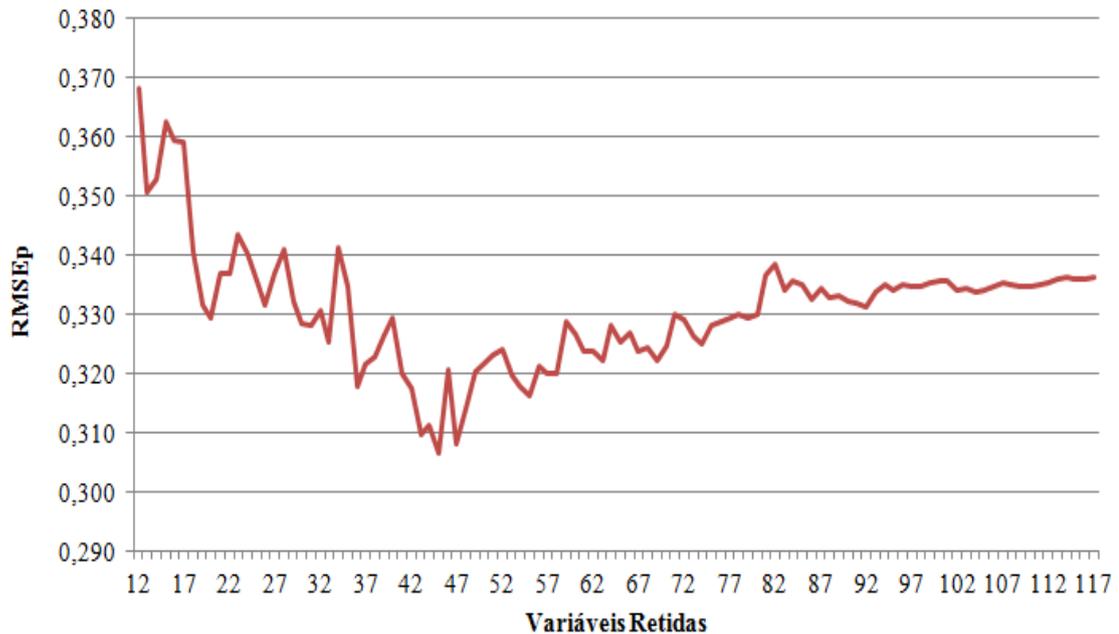
É interessante analisar o comportamento de variáveis altamente correlacionadas, como o subconjunto  $x_{206}$  a  $x_{210}$ , que apresentam coeficiente de correlação acima de 0,99. A variável  $x_{208}$  foi utilizada para calcular as variáveis de resposta  $y_{203}$  e  $y_{206}$ , com correlações positivas fortes ( $r = 0,87$  e  $0,73$ , respectivamente). Esperava-se, portanto, que os coeficientes de regressão  $b$  associando estas variáveis fossem maiores que os das demais variáveis, uma vez que os dados estão padronizados. Contudo, observa-se que, para  $y_{203}$ , o maior coeficiente de regressão é 0,490 para a variável  $x_{131}$ , também utilizada para o cálculo de  $y_{203}$ , mas com uma correlação menor ( $r = 0,42$ ) e, em seguida, as variáveis retidas desse subconjunto,  $x_{207}$ ,  $x_{208}$  e  $x_{209}$ , com coeficientes iguais a 0,225, 0,237 e 0,226, respectivamente.

No terceiro estágio, a variância explicada de  $Y$  pelo modelo chegou a 92,5% ( $R^2_Y = 0,925$ ), sendo retidas 45 variáveis independentes, conforme Figura 27: 13 do primeiro estágio, 11 do segundo e 21 do terceiro estágio. De acordo com a Tabela 12, os relacionamentos gerados contemplam 4 variáveis do primeiro estágio ( $x_{102}$ ,  $x_{114}$ ,  $x_{122}$  e  $x_{126}$ ), 3 do segundo ( $x_{205}$ ,  $x_{220}$  e  $x_{224}$ ) e 8 do terceiro ( $x_{304}$ ,  $x_{310}$ ,  $x_{315}$ ,  $x_{320}$ ,  $x_{321}$ ,  $x_{324}$ ,  $x_{326}$  e  $x_{335}$ ), sendo que alguns destes com correlações de pequenas a moderadas (entre 0,30 e 0,50).

Analisando as variáveis do primeiro estágio, além das 4 variáveis esperadas, outras 9 foram retidas:  $x_{106}$ ,  $x_{113}$ ,  $x_{119}$ ,  $x_{127}$ ,  $x_{128}$ ,  $x_{130}$ ,  $x_{134}$ ,  $x_{136}$  e  $x_{138}$ . Destas,  $x_{130}$  e  $x_{134}$  permaneceram devido à elevada correlação com  $y_{301}$  e as outras por conta de redundâncias apresentadas:  $x_{106}$  com  $x_{335}$  (correlação moderada,  $r = 0,47$ );  $x_{113}$  com  $x_{114}$  (correlação forte,  $r = 0,77$ );  $x_{119}$  com  $x_{122}$  (correlação forte,  $r = 0,75$ );  $x_{127}$  com  $x_{220}$  (correlação moderada,  $r = 0,69$ );  $x_{128}$  com  $x_{220}$  (correlação moderada,  $r = 0,61$ );  $x_{136}$  com  $x_{335}$  (correlação moderada,  $r = 0,45$ ) e  $x_{138}$  com  $x_{335}$  (correlação moderada,  $r = 0,40$ ).

Das variáveis do segundo estágio, permaneceram no modelo  $x_{204}$ ,  $x_{205}$ ,  $x_{211}$ ,  $x_{212}$ ,  $x_{214}$ ,  $x_{218}$ ,  $x_{219}$ ,  $x_{221}$ ,  $x_{223}$ ,  $x_{224}$  e  $x_{227}$ . Como esperado, as variáveis  $x_{205}$  e  $x_{224}$  ficaram retidas no modelo, porém a variável  $x_{220}$  foi substituída por  $x_{127}$  e  $x_{128}$  (correlações moderadas,  $r = 0,69$  e  $r = 0,61$ ), enquanto a redundância manteve outras ( $x_{204}$  com  $x_{331}$ ,  $r = -0,59$ ;  $x_{211}$  com  $x_{335}$ ,  $r = -0,50$ ;  $x_{219}$  com  $x_{321}$ ,  $r = 0,58$ ;  $x_{221}$  com  $x_{223}$ ,  $r = 0,64$  e  $x_{223}$  com  $x_{320}$ ,  $r = 0,43$ ) e além destas permaneceram as variáveis independentes  $x_{212}$ ,  $x_{214}$ ,  $x_{218}$  e  $x_{227}$  por conta de  $y_{301}$ .

Figura 27 – RMSEp em função do número de variáveis retidas no estágio 3 para o estudo simulado



Fonte: Elaborada pelo autor

Para o terceiro estágio, além das 8 variáveis que presumidamente seriam retidas, foram mantidas outras por conta de suas redundâncias:  $x_{301}$  com  $x_{126}$ ,  $r = 0,75$ ;  $x_{308}$ ,  $x_{311}$  e  $x_{312}$  com  $x_{310}$ ,  $r \geq 0,99$ ;  $x_{319}$  e  $x_{322}$  com  $x_{320}$ ,  $r = 0,90$ ;  $x_{325}$  com  $x_{315}$ ,  $r = 0,43$ ;  $x_{330}$  com  $x_{302}$ ,  $r = -0,78$ ;  $x_{336}$  com  $x_{335}$ ,  $r = 0,51$  e  $x_{338}$  com  $x_{326}$ ,  $r = 0,64$ ).

Ao comparar-se o percentual de variáveis retidas do estudo simulado (Tabela 14) com o caso real (ver Tabela 1), percebem-se índices semelhantes de redução de variáveis, sugerindo consistência do método em função do número de variáveis originais. Ainda há de se salientar que uma parcela significativa das retenções no estudo simulado ocorreu devido ao efeito de redundância.

Tabela 14 – Percentual de variáveis retidas por estágio, no estudo simulado

Estágio	Variáveis Independentes	Variáveis Retidas	% Variáveis Retidas
1	39	13	33,3
2	78	22	28,2
3	117	45	38,5

Fonte: Elaborada pelo autor

Após a seleção das variáveis, seguindo os passos do método proposto, gerou-se uma regressão PLS para cada estágio, sendo os coeficientes de regressão apresentados no Apêndice F.

Em seguida, foram definidos os limites de especificação e controle, e calculados os índices de capacidade do processo. Os limites de especificação determinados sempre eram menores do que os limites de controle, de maneira que todos os índices de capacidade mostrassem incapacidade de atendimento das especificações e poucas variáveis centradas (indicando que, na maioria das variáveis, há necessidade de alteração na média). O índice de importância foi definido como 1 para todas variáveis de resposta.

O modelo de programação quadrática foi construído (ver Apêndice H) e otimizado, obtendo-se as condições ótimas do processo para as variáveis independentes, conforme Tabela 15. Os valores das variáveis de resposta decorrentes de tais níveis são apresentados na Tabela 16.

Os resultados obtidos, conforme Tabela 16, mostram a expectativa de que a maioria dos valores nominais sejam alcançados, principalmente das variáveis de resposta do segundo e terceiro estágios; contudo a variável  $y_{104}$  tem valor nominal de 10 e o valor encontrado foi de 9,314. Supondo que seja uma importante característica de qualidade pode-se aumentar o índice de importância, penalizando o modelo por este desvio do valor nominal. Assim, realizou-se uma simulação considerando índice de importância 3 para esta variável e o valor esperado aumentou para 9,351 e considerando índice de importância 5, o valor esperado foi de 9,384. Observa-se que aumentar o índice de importância não garante a obtenção do valor nominal, porém diminui o desvio e pode ser uma estratégia a ser utilizada nessas situações.

Um modelo de regressão PLS completo, contemplando todas as variáveis e mesmo número de componentes em cada estágio (5, 11 e 12) também foi criado a título de comparação com o modelo proposto. Na Tabela 17 se percebe que o erro médio de previsão ( $RMSE_p$ ) é menor no modelo proposto em todos os estágios e os coeficientes de determinação são semelhantes.

Tabela 15– Condição ótima de trabalho para o estudo simulado

Variável Independente	Valor para ajuste	Variável Independente	Valor para ajuste
$x_{102}$	0,251	$x_{214}$	-1,011
$x_{106}$	-0,012	$x_{218}$	-1,117
$x_{108}$	0,481	$x_{219}$	-0,902
$x_{109}$	-0,776	$x_{220}$	-0,268
$x_{113}$	-0,145	$x_{221}$	1,373
$x_{114}$	-0,404	$x_{222}$	-0,414
$x_{118}$	-0,616	$x_{223}$	0,573
$x_{119}$	1,052	$x_{224}$	-0,872
$x_{121}$	-1,536	$x_{227}$	1,696
$x_{122}$	1,465	$x_{301}$	1,435
$x_{123}$	-1,537	$x_{304}$	0,177
$x_{125}$	0,184	$x_{308}$	0,665
$x_{126}$	1,170	$x_{310}$	0,654
$x_{127}$	-1,006	$x_{311}$	0,713
$x_{128}$	0,746	$x_{312}$	0,835
$x_{130}$	-0,854	$x_{315}$	-0,310
$x_{131}$	-0,662	$x_{319}$	0,995
$x_{134}$	0,916	$x_{320}$	0,494
$x_{135}$	-1,121	$x_{321}$	-1,124
$x_{136}$	0,553	$x_{322}$	1,512
$x_{137}$	-1,033	$x_{324}$	0,387
$x_{138}$	0,433	$x_{325}$	0,504
$x_{203}$	1,151	$x_{326}$	1,249
$x_{204}$	0,863	$x_{327}$	0,943
$x_{205}$	0,606	$x_{329}$	-0,310
$x_{207}$	0,925	$x_{330}$	-0,685
$x_{208}$	0,737	$x_{331}$	0,933
$x_{209}$	0,966	$x_{335}$	-0,753
$x_{211}$	0,111	$x_{336}$	-1,051
$x_{212}$	1,405	$x_{338}$	1,380

Fonte: Elaborada pelo autor

Tabela 16– Valores esperados nas variáveis de resposta na condição ótima para o estudo simulado

Variável de Resposta	Valor Nominal	Valor esperado	Diferença
y <sub>101</sub>	3,000	2,880	0,120
y <sub>102</sub>	1,750	1,772	0,022
y <sub>103</sub>	4,750	4,846	0,096
y <sub>104</sub>	10,000	9,314	0,686
y <sub>105</sub>	0,600	0,625	0,025
y <sub>106</sub>	2,000	1,931	0,069
y <sub>107</sub>	4,500	4,247	0,253
y <sub>108</sub>	-1,000	-1,239	0,239
y <sub>201</sub>	1,000	1,000	0,000
y <sub>202</sub>	3,500	3,500	0,000
y <sub>203</sub>	7,500	7,500	0,000
y <sub>204</sub>	4,000	4,000	0,000
y <sub>205</sub>	5,500	5,500	0,000
y <sub>206</sub>	12,000	12,000	0,000
y <sub>207</sub>	0,250	0,250	0,000
y <sub>208</sub>	1,750	1,750	0,000
y <sub>301</sub>	5,250	5,250	0,000
y <sub>302</sub>	0,000	0,000	0,000
y <sub>303</sub>	0,400	0,400	0,000
y <sub>304</sub>	7,800	7,800	0,000
y <sub>305</sub>	3,000	3,000	0,000
y <sub>306</sub>	11,000	11,000	0,000
y <sub>307</sub>	11,000	11,000	0,000
y <sub>308</sub>	9,000	9,000	0,000

Fonte: Elaborada pelo autor

Tabela 17 – Comparativo de resultados no estudo simulado entre o modelo completo e o modelo proposto

Indicador	Estágio	Modelo Completo	Modelo Proposto	Diferença
RMSEp	1	0,658	0,632	0,026
RMSEp	2	0,521	0,503	0,018
RMSEp	3	0,336	0,307	0,029
R <sup>2</sup> Y	1	58,7%	60,4%	1,7%
R <sup>2</sup> Y	2	77,9%	77,9%	0,0%
R <sup>2</sup> Y	3	91,8%	92,5%	0,7%
F. Objetivo		0,0052	0,0089	0,0037

Fonte: Elaborada pelo autor

Os resultados esperados para as variáveis de resposta no modelo completo estão na Tabela 18, na qual se observa que os valores esperados são muito próximos aos nominais e também aos do modelo proposto.

Tabela 18– Valores esperados nas variáveis de resposta na condição ótima para o estudo simulado

Variável de Resposta	Valor Nominal	Completo	Proposto	Diferença
y <sub>101</sub>	3,000	2,920	2,880	0,040
y <sub>102</sub>	1,750	1,775	1,772	0,003
y <sub>103</sub>	4,750	4,691	4,846	-0,155
y <sub>104</sub>	10,000	9,429	9,314	0,115
y <sub>105</sub>	0,600	0,586	0,625	-0,039
y <sub>106</sub>	2,000	2,006	1,931	0,076
y <sub>107</sub>	4,500	4,349	4,247	0,102
y <sub>108</sub>	-1,000	-1,223	-1,239	0,016
y <sub>201</sub>	1,000	1,003	1,000	0,003
y <sub>202</sub>	3,500	3,488	3,500	-0,012
y <sub>203</sub>	7,500	7,504	7,500	0,004
y <sub>204</sub>	4,000	3,992	4,000	-0,008
y <sub>205</sub>	5,500	5,494	5,500	-0,006
y <sub>206</sub>	12,000	11,983	12,000	-0,017
y <sub>207</sub>	0,250	0,253	0,250	0,003
y <sub>208</sub>	1,750	1,753	1,750	0,003
y <sub>301</sub>	5,250	5,248	5,250	-0,002
y <sub>302</sub>	0,000	0,000	0,000	0,000
y <sub>303</sub>	0,400	0,401	0,400	0,001
y <sub>304</sub>	7,800	7,803	7,800	0,003
y <sub>305</sub>	3,000	3,001	3,000	0,001
y <sub>306</sub>	11,000	10,998	11,000	-0,002
y <sub>307</sub>	11,000	11,010	11,000	0,010
y <sub>308</sub>	9,000	8,973	9,000	-0,027

Fonte: Elaborada pelo autor

Contudo, ao analisar-se os valores previstos para as variáveis independentes gerados pelo modelo completo, conforme Tabela 19, verifica-se que 117 variáveis independentes, a solução recomendada retém 84 variáveis no limite operacional (inferior ou superior) ou próximo dele, enquanto que no modelo gerado após a seleção de variáveis não há nenhuma das 60 variáveis retidas no limite ou mesmo próximo de um deles.

Tabela 19– Condição ótima de trabalho com o modelo completo para o estudo simulado

Variável Independente	Valor para ajuste	Variável Independente	Valor para ajuste	Variável Independente	Valor para ajuste
$x_{101}$	-2,350	$x_{201}$	2,366	$x_{301}$	1,801
$x_{102}$	-0,405	$x_{202}$	-1,159	$x_{302}$	2,503
$x_{103}$	2,313	$x_{203}$	2,135	$x_{303}$	2,544
$x_{104}$	-1,622	$x_{204}$	-0,755	$x_{304}$	0,965
$x_{105}$	-1,452	$x_{205}$	-2,380	$x_{305}$	-2,111
$x_{106}$	0,280	$x_{206}$	1,639	$x_{306}$	-2,336
$x_{107}$	-0,792	$x_{207}$	1,652	$x_{307}$	-1,959
$x_{108}$	-2,101	$x_{208}$	1,489	$x_{308}$	-0,193
$x_{109}$	-1,570	$x_{209}$	1,811	$x_{309}$	1,056
$x_{110}$	-2,078	$x_{210}$	1,852	$x_{310}$	0,707
$x_{111}$	-0,937	$x_{211}$	-1,434	$x_{311}$	1,237
$x_{112}$	-0,927	$x_{212}$	-1,870	$x_{312}$	1,369
$x_{113}$	-0,927	$x_{213}$	1,078	$x_{313}$	-2,039
$x_{114}$	0,964	$x_{214}$	-1,845	$x_{314}$	2,076
$x_{115}$	1,267	$x_{215}$	2,014	$x_{315}$	-2,392
$x_{116}$	-1,437	$x_{216}$	2,275	$x_{316}$	1,530
$x_{117}$	-1,281	$x_{217}$	1,454	$x_{317}$	-1,857
$x_{118}$	-1,508	$x_{218}$	1,720	$x_{318}$	1,826
$x_{119}$	-2,201	$x_{219}$	-1,367	$x_{319}$	2,267
$x_{120}$	2,125	$x_{220}$	-1,841	$x_{320}$	2,410
$x_{121}$	1,586	$x_{221}$	1,141	$x_{321}$	-1,707
$x_{122}$	1,052	$x_{222}$	1,854	$x_{322}$	1,774
$x_{123}$	-2,110	$x_{223}$	-0,251	$x_{323}$	-1,863
$x_{124}$	-0,456	$x_{224}$	-1,735	$x_{324}$	-0,230
$x_{125}$	-2,327	$x_{225}$	-0,937	$x_{325}$	2,339
$x_{126}$	-2,285	$x_{226}$	-1,568	$x_{326}$	1,712
$x_{127}$	2,097	$x_{227}$	2,158	$x_{327}$	1,402
$x_{128}$	1,277	$x_{228}$	1,983	$x_{328}$	-1,562
$x_{129}$	-1,039	$x_{229}$	1,721	$x_{329}$	-1,363
$x_{130}$	-1,030	$x_{230}$	-1,051	$x_{330}$	-1,652
$x_{131}$	1,238	$x_{231}$	-0,937	$x_{331}$	1,687
$x_{132}$	-1,392	$x_{232}$	-1,056	$x_{332}$	1,291
$x_{133}$	-1,240	$x_{233}$	0,580	$x_{333}$	1,561
$x_{134}$	-2,325	$x_{234}$	-0,892	$x_{334}$	-1,842
$x_{135}$	2,009	$x_{235}$	-0,309	$x_{335}$	-0,348
$x_{136}$	2,003	$x_{236}$	1,251	$x_{336}$	-2,708
$x_{137}$	2,266	$x_{237}$	2,079	$x_{337}$	-0,319
$x_{138}$	-1,539	$x_{238}$	2,364	$x_{338}$	1,768
$x_{139}$	1,776	$x_{239}$	2,595	$x_{339}$	-1,894

Fonte: Elaborada pelo autor

## 5 CONCLUSÕES

Neste capítulo estão as principais conclusões e contribuições deste trabalho, seguidas de sugestões para trabalhos futuros.

Este estudo teve como objetivo principal desenvolver um método para definição de condição ótima de operação para as variáveis mais relevantes de um processo contínuo multivariado e em múltiplos estágios. Para tanto, foi proposto um método de seleção de variáveis baseado na magnitude dos coeficientes da regressão PLS. Para cada estágio gerou-se um modelo de regressão PLS, utilizando-se do procedimento de validação cruzada para identificar-se o número de componentes. Após a definição do número de componentes gerou-se uma regressão PLS, calculou-se o erro médio de previsão ( $RMSE_p$ ) e aplicou-se um procedimento de classificar o impacto que cada variável independente gera nas variáveis de resposta. A variável independente com menor impacto foi eliminada e um novo modelo gerado com as variáveis remanescentes. Este procedimento de predição e eliminação foi sistematicamente realizado e, através do monitoramento do erro de previsão, identificou-se o modelo de regressão PLS de maior capacidade preditiva.

Tal sistemática de seleção apresentou bons resultados, descartando a maioria das variáveis independentes irrelevantes e aumentando a capacidade preditiva do modelo. Deve-se ainda salientar que o formato utilizado para construção das regressões PLS para cada estágio garante que o efeito cascata seja respeitado, o que não é garantido nas tradicionais abordagens com MBPLS propostas pela literatura.

Para a identificação das condições ótimas de operação do processo, utilizou-se um modelo de programação quadrática, o qual se apoiou nos coeficientes de regressão PLS para formatar o relacionamento entre as variáveis independentes selecionadas e as de resposta. Também neste aspecto a metodologia inova, pois leva em consideração a faixa definida pelos limites de especificação, o que não foi encontrado em outras abordagens.

Um estudo de caso com dados de uma empresa do setor avícola, ainda que com um número restrito de variáveis, mostrou a viabilidade de aplicação real da sistemática proposta, evidenciando ainda um ganho na capacidade de interpretação do processo. Na comparação com um modelo de regressão PLS completo, gerado com o mesmo número de componentes nos estágios e com todas as variáveis independentes, o modelo proposto foi superior em todos os estágios, com maior capacidade preditiva.

Realizou-se também um estudo simulado utilizando uma base de dados com maior número de variáveis independentes altamente correlacionadas. Os resultados corroboram a capacidade do método em reduzir o número de variáveis necessárias para predição das variáveis de resposta e subsequente otimização em processos multivariados e em múltiplos estágios e também de trabalhar em processos que apresentem um grande número de variáveis. Ao se comparar os resultados com um modelo de regressão PLS completo, o método proposto novamente apresentou maior capacidade preditiva e ainda determinou a condição ótima do processo com todas as variáveis independentes distantes dos limites operacionais: tal fato não ocorreu no modelo completo e, portanto, deixa dúvidas sobre a operacionalidade desta previsão. Este fato sugere que a aplicação do método de seleção reduz o risco de sobreajuste e, portanto, recomenda-se a sua utilização.

Conclui-se, assim, que a pesquisa trouxe contribuições importantes e que o objetivo principal de desenvolver uma sistemática de identificação da condição ótima de trabalho em processos multivariados e em múltiplos estágios foi alcançado, bem como todos os objetivos específicos. Em termos práticos, percebe-se ainda que o método proposto oferece grandes vantagens na gestão do processo, pois além de identificar as variáveis independentes que estão mais intimamente associadas às variáveis de resposta, reduz o número de variáveis a serem monitoradas, mitigando esforços de coleta e custos relacionados. Outra característica importante do método proposto em termos de sua originalidade é a capacidade de trabalhar em ambientes com múltiplas variáveis de resposta, o que o diferencia da maioria dos métodos de seleção reportados pela literatura.

Como sugestões para trabalhos futuros, sugere-se o aprofundamento de técnicas de seleção de variáveis voltadas a processos multiestágios, especialmente naqueles onde se verificam elevados níveis de correlação entre as variáveis de cada estágio. Outro fator citado na literatura e observado neste trabalho é a retenção de variáveis redundantes no processo de seleção, as quais podem dificultam a identificação das relações de causa e efeito, além de aumentarem o número de variáveis a ser monitoradas. Logo, sugere-se o desenvolvimento de sistemáticas de seleção de variáveis redundantes mais robustas.

Neste trabalho utilizou-se um indicador de importância das variáveis independentes ( $R_{qk}$ ) baseado estritamente nos coeficientes de regressão PLS. De tal forma, entende-se o desenvolvimento de novos índices de importância como uma potencial área a ser considerada.

Condição comum em processos industriais (principalmente nas indústrias químicas) e não contemplada neste trabalho, a ocorrência de ciclos aumenta a complexidade da modelagem para se encontrar as condições ótimas de operação. Desta forma, este é um cenário a ser estudado.

Na metodologia proposta, optou-se por uma abordagem de atualização das condições ótimas somente ao fim de cada processo. No entanto, em processos de longa duração, seria viável a atualização após cada estágio, o que permitirá incluir no modelo de otimização os valores das variáveis independentes dos estágios anteriores. Sugere-se a ampliação deste estudo nestas condições de atualização estágio a estágio.

Outra expansão deste estudo poderia incluir técnicas de regressão não linear, como INLR (*implicit nonlinear latent variable regression*), GIFI, ou Kernel-PLS em processos onde relacionamentos lineares entre variáveis independentes e de resposta não são suficientes.

Considerando-se que, mesmo com o procedimento de seleção de variáveis, há ocorrência de variáveis redundantes, estudos contemplando a análise de correlações parciais após cada procedimento de eliminação de variáveis independentes pode fornecer uma melhor clareza no impacto que cada variável provoca no processo e facilitando a tomada de decisão sobre manter ou não essa variável no modelo preditivo.

Por fim, as variáveis independentes e de resposta utilizadas neste trabalho são contínuas. Entende-se que a consideração de variáveis binárias, comumente observadas em processos industriais, poderia ser um interessante desdobramento para avaliação da eficiência das proposições desta tese e para desenvolvimentos futuros.

## 6 REFERÊNCIAS BIBLIOGRÁFICAS

ABDI, H. Partial least squares regression and projection on latent structure regression (PLS Regression). **Wiley Interdisciplinary Reviews: Computational Statistics**, v. 2, 2010. 97-106.

ABRAHAMSSON, C. et al. Comparison of different variable selection methods conducted on NIR transmission measurements on intact tablets. **Chemometrics and Intelligent Laboratory Systems**, v. 69, 2003. 3-12.

ANDERSEN, C. M.; BRO, R. Variable selection in regression—a tutorial. **Journal of Chemometrics**, v. 24, 2010. 728–737.

ANZANELLO, M. J.; ALBIN, S. L.; CHAOVALITWONGSE, W. A. Selecting the best variables for classifying production batches into two quality levels. **Chemometrics and Intelligent Laboratory Systems**, 97, 2009. 111-117.

AVIAGEN. Matrizes Ross: Objetivos de Desempenho, 2011. Disponível em: <[http://pt.aviagen.com/assets/Tech\\_Center/BB\\_Foreign\\_Language\\_Docs/Portuguese/Ross308ObjDes\\_empehnoJun2011.pdf](http://pt.aviagen.com/assets/Tech_Center/BB_Foreign_Language_Docs/Portuguese/Ross308ObjDes_empehnoJun2011.pdf)>. Acesso em: 20 maio 2014.

AVIAGEN. ROSS Grandparent Management Manual, 2011. Disponível em: <[www.aviagen.com](http://www.aviagen.com)>. Acesso em: 20 maio 2014.

AVISITE. Brasil: uma década na liderança das exportações mundiais de carne de frango. Disponível em: <<http://www.avisite.com.br/noticias/index.php?codnoticia=15072>>. Acesso em: 29 maio 2014.

BASSI, N. S. S.; SILVA, C. L. D.; SANTOYO, A. Inovação, pesquisa e desenvolvimento na agroindústria avícola brasileira. **Estudos Sociedade e Agricultura**, Rio de Janeiro, v. 21, 2013. 392-417.

BERSIMIS, S.; PSARAKIS, S.; PANARETOS, J. Multivariate statistical process control charts: an overview. **Quality and Reliability Engineering International**, v. 23, 2007. 517–543.

BORGES, F. H.; DALCOL, P. R. T. Indústrias de processo: Comparações e caracterizações. **Anais do XXII Encontro Nacional de Engenharia de Produção**, Curitiba, out. 2002. 1-9.

BURNHAM, A. J.; MACGREGOR, J. F.; VIVEROS, R. Latent variable multivariate regression modeling. **Chemometrics and Intelligent Laboratory Systems**, v. 48, 1999. 167-180.

CAMACHO, J.; PICÓ, J.; FERRER, A. Multi-phase analysis framework for handling batch process data. **Journal of Chemometrics**, v. 22, 2008. 632-643.

CHEN, J.; TSAI, R.-K. Development of MBPLS based control for serial operation processes. **Korean Journal of Chemical Engineering**, v.26, 2009. 935-945.

CHONG, I.-G.; JUN, C.-H. Performance of some variable selection methods when multicollinearity is present. **Chemometrics and Intelligent Laboratory Systems**, v. 78, 2005. 103-112.

COLIN, E. C. **Pesquisa Operacional: 170 Aplicações em Estratégia, Finanças, Logística, Produção, Marketins e Vendas**. Rio de Janeiro: LTC, 2011.

DILL, D. D.; SILVA, A. P. D.; LUVIELMO, M. D. M. Processamento de empanados: sistemas de cobertura. **Estudos Tecnológicos**, v.5, n. 1, jan/abr 2009. 33-49.

- DOSTÁL, Z. **Optimal Quadratic Programming Algorithms**. New York: Springer, 2009.
- ESPÍNDOLA, C. J. **A cadeia produtiva de frango de corte na América do Sul: considerações preliminares**. 12º Encontro de Geógrafos da América Latina. Montevideu: Uruguai. 2009.
- FLORES-CERRILLO, J.; MACGREGOR, J. F. Control of batch product quality by trajectory manipulation using latent variable models. **Journal of Process Control**, v. 14, 2004. 539-553.
- FORINA, M.; CASOLINO, C.; MILLAN, C. P. Iterative Predictor Weighting (IPW) PLS: A technique for the elimination of useless predictors in regression problems. **Journal of Chemometrics**, v. 13, 1999. 165–184.
- FRANK, I. E. Intermediate least squares regression method. **Chemometrics and Intelligent Laboratory Systems**, v. 1, 1987. 233-242.
- FRANK, I. E.; KOWALSKI, B. R. Prediction of wine quality and geographic origin from chemical measurements by partial least-squares regression modeling. **Analytica Chimica Acta**, v. 162, 1984. 241-251.
- GARCÍA-MUÑOZ, S. et al. Troubleshooting of an industrial batch process using multivariate methods. **Industrial & Engineering Chemistry Research**, v. 42, 2003. 3592-3601.
- GARTHWAITE, P. H. An interpretation of partial least squares. **Journal of the American Statistical Association**, v. 89, 1994. 122-127.
- GAUCHI, J.-P.; CHAGNON, P. Comparison of selection methods of explanatory variables in PLS regression with application to manufacturing process data. **Chemometrics and Intelligent Laboratory Systems**, v. 58, 2001. 171-193.
- GELADI, P.; KOWALSKI, B. R. Partial least-squares regression: a tutorial. **Analytica Chimica Acta**, v. 185, 1986. 1-17.
- GERHARDT, T. E. E. S. D. T. **Método de Pesquisa**. 1. ed. Porto Alegre: Editora da UFRGS, 2009.
- GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2010.
- GILL, P. E.; WONG, E. Sequential Quadratic Programming Methods. In: LEE, J.; LEYFFER, S. **Mixed Integer Nonlinear Programming**. New York: Springer, 2012.
- GUYON, I.; ELISSEEFF, A. An introduction to variable and feature selection. **Journal of Machine Learning Research**, v. 3, 2003. 1157-1182.
- HAIR, J. J. F. et al. **Fundamentos de métodos de pesquisa em administração**. Porto Alegre: Bookman, 2005.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data mining, Inference and Prediction**. 2. ed. New York: Springer, 2009.
- HAWKINS, D. M. Regression adjustment for variables in multivariate quality control. **Journal of Quality Technology**, v. 25, n. 3, 1993. 170-182.
- HILLIER, F. S.; LIEBERMANN, G. J. **Introdução à Pesquisa Operacional**. 9a. ed. Porto Alegre: AMGH, 2013.
- HÖSKULDSSON, A. PLS Regression Methods. **Journal of Chemometrics**, v. 2, 1988. 211-228.

HÖSKULDSSON, A. Variable and subset selection in PLS regression. **Chemometrics and Intelligent Laboratory Systems**, v. 55, 2001. 23-38.

HYBRID. Performance Goals: Parent Stock Males, 2011. Disponível em: <[http://www.hybridturkeys.com/~media/Files/Hybrid/Converter/PS\\_Converter\\_Male\\_2011.pdf](http://www.hybridturkeys.com/~media/Files/Hybrid/Converter/PS_Converter_Male_2011.pdf)>. Acesso em: 20 maio 2014.

IBGE. Indicadores IBGE: Estatística de Produção Pecuária, p. 74, mar. 2014.

JACKSON, J. E. **A user's guide to principal components**. New York: John Wiley, 1991. 596 p.

JAECKLE, C. M.; MACGREGOR, J. F. Product Design through Multivariate Statistical Analysis of Process Data. **AIChE Journal**, v. 44, Maio 1998. 1105-1118.

JAECKLE, C.; MACGREGOR, J. F. Industrial applications of product design through the inversion of latent variable models. **Chemometrics and Intelligent Laboratory Systems**, v.50, 2000. 199-210.

JEARKPAPORN, D. et al. Process monitoring for mean shifts for multiple stage processes. **International Journal of Production Research**, v. 45, n. 23, 2007. 5547–5570.

KETTANEH, N.; BERGLUND, A.; WOLD, S. PCA and PLS with very large data sets. **Computational Statistics & Data Analysis**, v. 48, 2005. 69-85.

KIM, J.; HUANG, Q.; SHI, J. Latent variable based key process variable identification and process monitoring. **Journal of Manufacturing Systems**, 26, 2007. 53–61.

KOURTI, T. Abnormal situation detection, three-way data and projection methods; robust data archiving and modeling for industrial applications. **Annual Reviews in Control**, v. 27, 2003. 131–139.

KOURTI, T. Multivariate dynamic data modeling for analysis and statistical process control of batch processes, start-ups and grade transitions. **Journal of Chemometrics**, v. 17, 2003. 93-109.

KOURTI, T. Application of latent variable methods to process control and multivariate statistical process control in industry. **International Journal of Adaptive Control and Signal Processing**, v. 19, 2005. 213–246.

KOURTI, T.; MACGREGOR, J. F. Process analysis, monitoring and diagnosis, using multivariate projection methods. **Chemometrics and Intelligent Laboratory Systems**, v. 28, 1995. 3-21.

LACHTERMACHER, G. **Pesquisa Operacional na tomada de decisões**. 4a. ed. ed. São Paulo: Pearson Prentice Hall, 2009.

LATTIN, J. et al. **Análise de Dados Multivariados**. São Paulo: Cengage Learning, 2011.

LAZRAQ, A.; CÉROUX, R.; GAUCHI, J.-P. Selecting both latent and explanatory variables in the PLS1 regression model. **Chemometrics and Intelligent Laboratory Systems**, v. 66, 2003. 117-126.

LAZZARI, M. R. Avicultura de corte no Brasil: uma comparação entre as Regiões Sul e Centro-Oeste. **Indicadores Econômicos FEE**, Porto Alegre, v. 31, n. 4, 2004. 259-290.

LEARDI, R.; SEASHOLTZ, M. B.; PELL, R. J. Variable selection for multivariate calibration using a genetic algorithm: prediction of additive concentrations in polymer films from Fourier transform-infrared spectral data. **Analytica Chimica Acta**, v. 461, 2002. 189-200.

- LIU, Z. et al. Modeling and optimization of a tablet manufacturing line. **Journal of Pharmaceutical Innovation**, v. 6, 2011. 170-180.
- LIU, Z. et al. Scale-up of a Pharmaceutical Roller Compaction Process Using a Joint-Y Partial Least Squares Model. **Industrial & Engineering Chemistry Research**, v.50, 2011. 10696–10706.
- LUENBERGER, D. G.; YE, Y. **Linear and nonlinear programming**. 3. ed. New York: Springer, 2008.
- MACGREGOR, J. F. et al. Process monitoring and diagnosis by multiblock PLS methods. **AIChE Journal**, v.40, n. 5, maio 1994. 826-838.
- MACGREGOR, J. F. et al. Data-based latent variable methods for process analysis, monitoring and control. **Computers and Chemical Engineering**, v. 29, 2005. 1217-1223.
- MACGREGOR, J.; CINAR, A. Monitoring, fault diagnosis, fault tolerant. **Computers & Chemical Engineering**, v. 47, n. 20, Dez. 2012. 111–120.
- MANDROLI, S. S.; K., S. A.; DING, Y. A survey of inspection strategy and sensor distribution studies in discrete-part manufacturing processes. **IIE Transactions**, v. 38, 2006. 309-328.
- MCCARL, B. A.; MOSKOWITZ, H.; FURTAN, H. Quadratic Programming Applications. **The International Journal of Management Science**, v. 5, 1977. 43-55.
- MEHMOOD, T. et al. A review of variable selection methods in Partial Least Squares Regression. **Chemometrics and Intelligent Laboratory Systems**, v. 118, 2012. 62-69.
- MUTEKI, K.; MACGREGOR, J. F.; UEDA, T. Rapid development of new polymer blends: The optimal selection of materials of blend ratios. **Industrial & Engineering Chemistry Research**, v. 45, 2006. 4653-4660.
- NOCEDAL, J.; WRIGHT, S. J. **Numerical Optimization**. 2a. ed. New York: Springer, 2006.
- NOMIKOS, P.; MACGREGOR, J. F. Monitoring batch process using multiway component principal analysis. **AIChE Journal**, v.40, n. 8, 1994. 1361-1375.
- NOMIKOS, P.; MACGREGOR, J. F. Multi-way partial least squares in monitoring batch processes. **Chemometrics and Intelligent Laboratory Systems**, v. 30, 1995. 97-108.
- NØRGAARD, L. et al. Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. **Applied Spectroscopy**, v. 54, 2000. 413-419.
- OAKLAND, J. S. **Statistical Process Control**. 6a. ed. Oxford: Elsevier, 2008.
- PEDRO, A. M. K.; FERREIRA, M. M. C. Simultaneously calibrating solids, sugars and acidity of tomato products using PLS2 and NIR spectroscopy. **Analytica Chimica Acta**, 595, 2007. 221–227.
- PIERNA, J. A. F. et al. A Backward Variable Selection method for PLS regression (BVSPLS). **Analytica Chimica Acta**, v. 642, 2009. 89-93.
- QIN, S. J. Survey on data-driven industrial process monitoring and diagnosis. **Annual Reviews in Control**, v. 36, 2012. 220–234.
- RAICH, A.; ÇINAR, A. Statistical process monitoring and disturbance diagnosis in multivariable continuous processes. **AIChE Journal**, v. 42, 1996. 995-1009.

REINIKAINEN, S.-P.; HÖSKULDSSON, A. Multivariate statistical analysis of a multi-step industrial processes. **Analytica Chimica Acta**, v. 595, 2007. 248-256.

ROSIPAL, R.; KRÄMER, N. **Overview and recent advances in partial least squares**. Proceedings of the 2005 international conference on Subspace, Latent Structure and Feature Selection. Bohinj, Slovenia: Springer. 2005. p. 34-51.

RURALBR. Ubabef projeta alta para produção e exportação de carne de frango em 2014. Disponível em: <<http://pecuaria.ruralbr.com.br/noticia/2014/01/ubabef-projeta-alta-para-producao-e-exportacao-de-carne-de-frango-em-2014-4391589.html>>. Acesso em: 24 abr. 2014.

SAEYS, Y.; INZA, I.; LARRAÑAGA, P. A review of feature selection techniques in bioinformatics. **Bioinformatics**, v. 23, 2007. 2507-2517.

SANTINI, G. A. **Dinâmica tecnológica na cadeia de frangos de corte no Brasil: análise dos segmentos de insumos e processamento**. Tese de Doutorado. Universidade Federal de São Carlos. São Carlos. 2006.

SARCINELLI, M. F.; VENTURINI, K. S.; SILVA, L. C. D. **Processamento de Carne de Frango**. UFES. Vitória, p. 1-7. 2007.

SEAB/PR. Análise de Conjuntura Agropecuária: 2012-2013. **Secretaria de Agricultura e Abastecimento do Paraná**, 2014. Disponível em: <[http://www.agricultura.pr.gov.br/arquivos/File/deral/Prognosticos/peru\\_2012\\_13.pdf](http://www.agricultura.pr.gov.br/arquivos/File/deral/Prognosticos/peru_2012_13.pdf)>. Acesso em: 31 maio 2014.

SHI, J.; ZHOU, S. Quality control and improvement for multistage systems: a survey. **IEEE Transactions**, v. 41, 2009. 744-753.

SIQUEIRA, L. G. P. **Controle Estatístico do Processo**. São Paulo: Pioneira, 1997.

SKAGERBERG, B.; MACGREGOR, J. F.; KIPARISSIDES, C. Multivariate data analysis applied to low-density polyethylene reactors. **Chemometrics and Intelligent Laboratory Systems**, v. 14, 1992. 341-356.

STEVENSON, W. J. **Estatística Aplicada à Administração**. São Paulo: Harper & Row do Brasil, 1981.

TOMBA, E.; BAROLO, M.; GARCÍA-MUÑOZ, S. General Framework for Latent Variable Model Inversion for the Design and Manufacturing of New Products. **Industrial & Engineering Chemistry Research**, v. 51, 2012. 12886–12900.

UBABEF. Estatísticas de Frango. Disponível em: <<http://www.ubabef.com.br/estatisticas/frango>>. Acesso em: 24 maio 2014.

UBABEF. Relatório Anual 2013. Disponível em: <<http://www.ubabef.com.br/files/publicacoes/732e67e684103de4a2117dda9ddd280a.pdf>>. Acesso em: 30 maio 2014.

UBABEF. Relatório Anual 2014. Disponível em: <<http://www.ubabef.com.br/files/publicacoes/8ca705e70f0cb110ae3aed67d29c8842.pdf>>. Acesso em: 05 jun. 2014.

ÜNDEY, C.; ÇINAR, A. Statistical monitoring of multistage, multiphase batch processes. **IEEE Control System Magazine**, v.22, 2002. 40-52.

- VANDEGINSTE, D. G. M. et al. Handbook of Chemometrics and Qualimetrics. Part B. 1. ed. Amsterdam: Elsevier Science, v. 20B, 1998. p. 349-370.
- VEGRO, C. L. R.; ROCHA, M. B. Expectativas tecnológicas para o segmento de carnes de aves e suínos. **Informações Econômicas**, São Paulo, v.37, n. 5, maio 2007. 15-28.
- VENKATASUBRAMANIAN, V. et al. A review of process fault detection and diagnosis- Part III: Process history based methods. **Computers and Chemical Engineering**, v. 27, 2003. 327-346.
- VINZI, V. E. et al. **Handbook of Partial Least Squares. Concepts, Methods and Applications**. Berlin: Springer, 2010.
- VOILÀ, M.; TRICHES, D. **A cadeia de carne de frango: uma análise dos mercados brasileiro e mundial de 2002 a 2010**. IPES / UCS. Caxias do Sul, p. 2-24. 2013.
- WESTERHUIS, J. A.; COENEGRACHT, P. M. J. Multivariate modelling of the pharmaceutical two-step process of wet granulation and tableting with multi-block partial least squares. **Journal of Chemometrics**, v. 11, 1997. 379–392.
- WOLD, S.; SJÖSTRÖM, M.; ERIKSSON, L. PLS-regression: a basic tool of chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v. 58, 2001. 109–130.
- YACOUB, F.; MACGREGOR, J. F. Product optimization and control in the latent variable space of nonlinear PLS models. **Chemometrics and Intelligent Laboratory Systems**, v. 70, 2004. 63-74.
- YAO, Y.; GAO, F. A survey on multistage/multiphase statistical modeling methods for batch processes. **Annual Reviews in Control**, v. 33, 2009. 172-183.
- ZARZO, M.; FERRER, A. Batch process diagnosis: PLS with variable selection versus block-wise PCR. **Chemometrics and Intelligent Laboratory Systems**, v. 73, 2004. 15-27.
- ZHANG, Y.; TENG, Y.; ZHANG, Y. Complex process quality prediction using modified kernel partial least squares. **Chemical Engineering Science**, v. 65, 2010. 2153-2158.
- ZIMMER, J.; ANZANELLO, M. J. Um novo método para seleção de variáveis preditivas com base em índices de importância. **Produção**, São Paulo, v. 24, 2014. 84-93.

**APÊNDICES**

## Apêndice A – Matrizes de correlação por estágio

<b>Matriz dos coeficientes de correlação das variáveis independentes e de resposta do estágio 1</b>																	
	X101	X102	X103	X104	X105	X106	X107	X108	X109	X110	X111	Y101	Y102	Y103	Y104	Y105	Y106
X101	<b>1</b>	0,13	<b>0,79</b>	<b>0,56</b>	0,01	-0,14	0,04	0,09	0,16	0,06	-0,07	<b>0,74</b>	<b>0,85</b>	0,15	0,02	<b>0,85</b>	0,18
X102	0,13	<b>1</b>	-0,26	0,16	0,29	-0,09	0,03	-0,12	0,08	-0,06	0,06	-0,06	0,14	<b>0,64</b>	-0,08	0,04	0,27
X103	<b>0,79</b>	-0,26	<b>1</b>	<b>0,49</b>	-0,19	-0,05	-0,03	0,06	0,13	0,05	-0,09	<b>0,87</b>	<b>0,67</b>	-0,32	0,03	<b>0,75</b>	-0,02
X104	<b>0,56</b>	0,16	<b>0,49</b>	<b>1</b>	-0,08	-0,11	-0,03	-0,05	-0,11	-0,02	-0,02	<b>0,72</b>	<b>0,28</b>	0,06	-0,05	<b>0,22</b>	<b>0,82</b>
X105	0,01	0,29	-0,19	-0,08	<b>1</b>	-0,17	0,09	-0,05	0,14	0,00	0,06	-0,14	0,08	<b>0,31</b>	0,03	0,03	0,00
X106	-0,14	-0,09	-0,05	-0,11	-0,17	<b>1</b>	-0,04	0,08	-0,06	-0,19	0,08	-0,07	-0,09	-0,07	-0,02	-0,08	-0,09
X107	0,04	0,03	-0,03	-0,03	0,09	-0,04	<b>1</b>	0,39	0,22	-0,05	0,06	-0,05	0,07	0,09	<b>0,98</b>	0,06	-0,02
X108	0,09	-0,12	0,06	-0,05	-0,05	0,08	0,39	<b>1</b>	0,12	0,03	-0,02	0,02	0,11	0,03	0,40	0,12	-0,08
X109	0,16	0,08	0,13	-0,11	0,14	-0,06	0,22	0,12	<b>1</b>	0,01	-0,07	0,12	0,18	0,12	0,19	<b>0,21</b>	-0,27
X110	0,06	-0,06	0,05	-0,02	0,00	-0,19	-0,05	0,03	0,01	<b>1</b>	0,01	0,04	0,08	0,04	-0,06	0,08	-0,05
X111	-0,07	0,06	-0,09	-0,02	0,06	0,08	0,06	-0,02	-0,07	0,01	<b>1</b>	0,00	-0,09	0,09	0,04	-0,09	0,00
Y101	<b>0,74</b>	-0,06	<b>0,87</b>	<b>0,72</b>	-0,14	-0,07	-0,05	0,02	0,12	0,04	0,00	<b>1</b>	<b>0,53</b>	-0,05	-0,05	<b>0,58</b>	0,19
Y102	<b>0,85</b>	0,14	<b>0,67</b>	<b>0,28</b>	0,08	-0,09	0,07	0,11	0,18	0,08	-0,09	<b>0,53</b>	<b>1</b>	0,33	0,03	<b>0,98</b>	-0,03
Y103	0,15	<b>0,64</b>	-0,32	0,06	0,31	-0,07	0,09	0,03	0,12	0,04	0,09	-0,05	0,33	<b>1</b>	-0,07	0,21	0,14
Y104	0,02	-0,08	0,03	-0,05	0,03	-0,02	<b>0,98</b>	0,40	0,19	-0,06	0,04	-0,05	0,03	-0,07	<b>1</b>	0,04	-0,05
Y105	<b>0,85</b>	0,04	<b>0,75</b>	<b>0,22</b>	0,03	-0,08	0,06	0,12	0,21	0,08	-0,09	<b>0,58</b>	<b>0,98</b>	0,21	0,04	<b>1</b>	-0,16
Y106	0,18	0,27	-0,02	<b>0,82</b>	0,00	-0,09	-0,02	-0,08	-0,27	-0,05	0,00	0,19	-0,03	0,14	-0,05	-0,16	<b>1</b>

Matriz dos coeficientes de correlação das variáveis independentes e de resposta do estágio 2

	X101	X102	X103	X104	X105	X106	X107	X108	X109	X110	X111	X201	X202	X203	X204	X205	X206	X207	Y201	Y202	Y203	Y204	Y205	Y206
X101	1	0,13	0,79	0,56	0,01	-0,14	0,04	0,09	0,16	0,06	-0,07	-0,06	-0,02	-0,03	-0,59	-0,67	0,02	0,01	0,74	0,84	-0,18	-0,01	0,02	0,52
X102	0,13	1	-0,26	0,16	0,29	-0,09	0,03	-0,12	0,08	-0,06	0,06	-0,16	0,26	0,05	-0,06	0,05	0,15	-0,01	-0,05	0,13	0,08	0,16	0,18	-0,14
X103	0,79	-0,26	1	0,49	-0,19	-0,05	-0,03	0,06	0,13	0,05	-0,09	0,00	-0,15	-0,09	-0,57	-0,79	-0,02	0,02	0,87	0,65	-0,20	-0,19	-0,05	0,67
X104	0,56	0,16	0,49	1	-0,08	-0,11	-0,03	-0,05	-0,11	-0,02	-0,02	-0,04	-0,05	-0,05	-0,50	-0,58	0,09	0,08	0,71	0,27	-0,04	-0,03	0,00	0,53
X105	0,01	0,29	-0,19	-0,08	1	-0,17	0,09	-0,05	0,14	0,00	0,06	-0,19	0,33	0,19	-0,01	0,04	0,05	0,07	-0,13	0,07	-0,02	0,27	0,32	-0,09
X106	-0,14	-0,09	-0,05	-0,11	-0,17	1	-0,04	0,08	-0,06	-0,19	0,08	0,11	-0,09	-0,05	0,14	0,16	-0,01	-0,03	-0,07	-0,09	0,12	-0,12	-0,11	-0,11
X107	0,04	0,03	-0,03	-0,03	0,09	-0,04	1	0,39	0,22	-0,05	0,06	-0,16	0,21	0,05	0,04	0,17	0,01	-0,01	-0,05	0,06	0,03	0,16	0,15	-0,04
X108	0,09	-0,12	0,06	-0,05	-0,05	0,08	0,39	1	0,12	0,03	-0,02	-0,10	-0,02	0,02	0,04	-0,02	0,00	-0,11	0,02	0,10	0,14	-0,01	0,09	0,09
X109	0,16	0,08	0,13	-0,11	0,14	-0,06	0,22	0,12	1	0,01	-0,07	-0,05	0,03	0,05	-0,08	-0,08	0,09	0,04	0,13	0,17	-0,04	-0,01	0,10	0,09
X110	0,06	-0,06	0,05	-0,02	0,00	-0,19	-0,05	0,03	0,01	1	0,01	0,03	-0,07	-0,04	-0,13	-0,02	0,03	0,10	0,04	0,07	-0,15	0,11	-0,04	0,10
X111	-0,07	0,06	-0,09	-0,02	0,06	0,08	0,06	-0,02	-0,07	0,01	1	0,00	-0,03	-0,03	0,05	-0,04	-0,12	0,00	0,00	-0,09	0,10	0,08	0,00	-0,04
X201	-0,06	-0,16	0,00	-0,04	-0,19	0,11	-0,16	-0,10	-0,05	0,03	0,00	1	-0,44	-0,53	-0,01	-0,03	-0,10	-0,05	0,03	-0,11	-0,10	-0,19	-0,78	-0,03
X202	-0,02	0,26	-0,15	-0,05	0,33	-0,09	0,21	-0,02	0,03	-0,07	-0,03	-0,44	1	0,32	0,07	0,14	-0,02	0,19	-0,14	0,04	0,15	0,18	0,61	-0,11
X203	-0,03	0,05	-0,09	-0,05	0,19	-0,05	0,05	0,02	0,05	-0,04	-0,03	-0,53	0,32	1	0,09	0,12	0,01	-0,01	-0,12	-0,05	0,02	0,11	0,63	-0,09
X204	-0,59	-0,06	-0,57	-0,50	-0,01	0,14	0,04	0,04	-0,08	-0,13	0,05	-0,01	0,07	0,09	1	0,65	-0,04	0,10	-0,70	-0,39	0,19	-0,19	0,07	-0,45
X205	-0,67	0,05	-0,79	-0,58	0,04	0,16	0,17	-0,02	-0,08	-0,02	-0,04	-0,03	0,14	0,12	0,65	1	-0,09	0,06	-0,89	-0,51	0,30	0,09	0,04	-0,67
X206	0,02	0,15	-0,02	0,09	0,05	-0,01	0,01	0,00	0,09	0,03	-0,12	-0,10	-0,02	0,01	-0,04	-0,09	1	0,00	0,06	0,01	-0,01	-0,01	0,08	-0,02
X207	0,01	-0,01	0,02	0,08	0,07	-0,03	-0,01	-0,11	0,04	0,10	0,00	-0,05	0,19	-0,01	0,10	0,06	0,00	1	-0,01	0,00	0,11	-0,10	0,12	0,07
Y201	0,74	-0,05	0,87	0,71	-0,13	-0,07	-0,05	0,02	0,13	0,04	0,00	0,03	-0,14	-0,12	-0,70	-0,89	0,06	-0,01	1	0,50	-0,18	-0,15	-0,07	0,72
Y202	0,84	0,13	0,65	0,27	0,07	-0,09	0,06	0,10	0,17	0,07	-0,09	-0,11	0,04	-0,05	-0,39	-0,51	0,01	0,00	0,50	1	-0,12	0,10	0,06	0,32
Y203	-0,18	0,08	-0,20	-0,04	-0,02	0,12	0,03	0,14	-0,04	-0,15	0,10	-0,10	0,15	0,02	0,19	0,30	-0,01	0,11	-0,18	-0,12	1	0,03	0,11	-0,05
Y204	-0,01	0,16	-0,19	-0,03	0,27	-0,12	0,16	-0,01	-0,01	0,11	0,08	-0,19	0,18	0,11	-0,19	0,09	-0,01	-0,10	-0,15	0,10	0,03	1	0,17	-0,14
Y205	0,02	0,18	-0,05	0,00	0,32	-0,11	0,15	0,09	0,10	-0,04	0,00	-0,78	0,61	0,63	0,07	0,04	0,08	0,12	-0,07	0,06	0,11	0,17	1	0,00
Y206	0,52	-0,14	0,67	0,53	-0,09	-0,11	-0,04	0,09	0,09	0,10	-0,04	-0,03	-0,11	-0,09	-0,45	-0,67	-0,02	0,07	0,72	0,32	-0,05	-0,14	0,00	1

Matriz dos coeficientes de correlação das variáveis independentes e de resposta do estágio 3

	X101	X102	X103	X104	X105	X106	X107	X108	X109	X110	X111	X201	X202	X203	X204	X205	X206	X207	X301	X302	X303	X304	X305	X306	X307	X308	Y301	Y302	Y303	Y304	Y305	Y306	Y307	Y308	Y309	Y310
X101	1	0,13	0,79	0,56	0,01	-0,14	0,04	0,09	0,16	0,06	-0,07	-0,06	-0,02	-0,03	-0,59	-0,67	0,02	0,01	0,68	0,70	0,51	0,51	0,60	-0,58	0,02	0,02	-0,32	0,27	-0,02	0,03	0,08	-0,30	-0,49	-0,22	0,81	0,00
X102	0,13	1	-0,26	0,16	0,29	-0,09	0,03	-0,12	0,08	-0,06	0,06	-0,16	0,26	0,05	-0,06	0,05	0,15	-0,01	-0,04	-0,02	-0,02	-0,03	-0,11	-0,11	0,08	-0,05	-0,09	0,06	0,08	0,02	-0,01	-0,12	-0,02	0,03	0,12	0,17
X103	0,79	-0,26	1	0,49	-0,19	-0,05	-0,03	0,06	0,13	0,05	-0,09	0,00	-0,15	-0,09	-0,57	-0,79	-0,02	0,02	0,80	0,80	0,62	0,55	0,74	-0,54	-0,05	0,06	-0,28	0,24	-0,02	0,05	0,10	-0,27	-0,53	-0,22	0,63	-0,05
X104	0,56	0,16	0,49	1	-0,08	-0,11	-0,03	-0,05	-0,11	-0,02	-0,02	-0,04	-0,05	-0,05	-0,50	-0,58	0,09	0,08	0,63	0,68	0,53	0,44	0,55	-0,49	-0,03	0,08	-0,25	0,32	-0,04	-0,02	0,01	-0,29	-0,50	-0,15	0,25	0,00
X105	0,01	0,29	-0,19	-0,08	1	-0,17	0,09	-0,05	0,14	0,00	0,06	-0,19	0,33	0,19	-0,01	0,04	0,05	0,07	-0,10	-0,14	-0,11	-0,13	-0,15	-0,11	0,14	-0,06	0,07	-0,03	0,08	0,07	0,05	-0,09	0,14	0,01	0,05	0,32
X106	-0,14	-0,09	-0,05	-0,11	-0,17	1	-0,04	0,08	-0,06	-0,19	0,08	0,11	-0,09	-0,05	0,14	0,16	-0,01	-0,03	-0,07	-0,06	-0,06	-0,03	-0,05	0,18	-0,19	0,05	0,02	0,01	-0,12	0,02	-0,06	0,08	0,03	0,01	-0,08	-0,10
X107	0,04	0,03	-0,03	-0,03	0,09	-0,04	1	0,39	0,22	-0,05	0,06	-0,16	0,21	0,05	0,04	0,17	0,01	-0,01	-0,01	-0,07	-0,03	-0,02	0,06	-0,01	-0,13	0,05	0,08	-0,15	0,19	0,05	0,10	0,17	0,13	0,02	0,05	0,16
X108	0,09	-0,12	0,06	-0,05	-0,05	0,08	0,39	1	0,12	0,03	-0,02	-0,10	-0,02	0,02	0,04	-0,02	0,00	-0,11	0,07	-0,02	-0,02	0,03	0,04	0,01	-0,10	0,13	-0,03	-0,11	-0,02	0,23	0,00	0,05	-0,02	-0,05	0,10	0,09
X109	0,16	0,08	0,13	-0,11	0,14	-0,06	0,22	0,12	1	0,01	-0,07	-0,05	0,03	0,05	-0,08	-0,08	0,09	0,04	0,17	0,08	0,07	-0,09	0,20	-0,12	0,11	-0,11	0,07	-0,03	0,19	0,09	0,15	0,08	0,05	-0,07	0,11	0,12
X110	0,06	-0,06	0,05	-0,02	0,00	-0,19	-0,05	0,03	0,01	1	0,01	0,03	-0,07	-0,04	-0,13	-0,02	0,03	0,10	0,06	0,00	0,04	-0,10	0,14	-0,10	0,00	-0,05	0,05	-0,02	-0,15	0,03	0,07	0,10	0,12	0,00	0,05	-0,04
X111	-0,07	0,06	-0,09	-0,02	0,06	0,08	0,06	-0,02	-0,07	0,01	1	0,00	-0,03	-0,03	0,05	-0,04	-0,12	0,00	-0,01	0,01	0,02	0,02	0,04	0,07	-0,07	-0,01	0,02	-0,02	0,07	-0,03	0,00	-0,03	0,11	0,03	-0,09	0,00
X201	-0,06	-0,16	0,00	-0,04	-0,19	0,11	-0,16	-0,10	-0,05	0,03	0,00	1	-0,44	-0,53	-0,01	-0,03	-0,10	-0,05	-0,06	0,07	0,07	0,01	0,03	0,22	0,07	-0,01	0,03	0,15	-0,03	-0,08	-0,27	0,03	-0,06	0,07	-0,09	-0,75
X202	-0,02	0,26	-0,15	-0,05	0,33	-0,09	0,21	-0,02	0,03	-0,07	-0,03	-0,44	1	0,32	0,07	0,14	-0,02	0,19	-0,06	-0,18	-0,06	-0,08	-0,12	-0,12	-0,06	0,05	-0,03	-0,21	0,10	0,03	0,24	0,01	0,16	0,09	0,05	0,59
X203	-0,03	0,05	-0,09	-0,05	0,19	-0,05	0,05	0,02	0,05	-0,04	-0,03	-0,53	0,32	1	0,09	0,12	0,01	-0,01	-0,09	-0,09	-0,23	0,03	-0,16	-0,10	-0,06	0,04	0,08	-0,07	0,07	-0,05	0,10	-0,04	0,04	-0,16	-0,05	0,57
X204	-0,59	-0,06	-0,57	-0,50	-0,01	0,14	0,04	0,04	-0,08	-0,13	0,05	-0,01	0,07	0,09	1	0,65	-0,04	0,10	-0,64	-0,67	-0,45	-0,47	-0,59	0,95	-0,02	-0,01	0,24	-0,29	0,03	-0,04	-0,07	0,27	0,44	0,26	-0,35	0,08
X205	-0,67	0,05	-0,79	-0,58	0,04	0,16	0,17	-0,02	-0,08	-0,02	-0,04	-0,03	0,14	0,12	0,65	1	-0,09	0,06	-0,80	-0,84	-0,65	-0,57	-0,70	0,62	0,04	0,02	0,27	-0,33	0,03	-0,05	-0,02	0,39	0,58	0,20	-0,48	0,04
X206	0,02	0,15	-0,02	0,09	0,05	-0,01	0,01	0,00	0,09	0,03	-0,12	-0,10	-0,02	0,01	-0,04	-0,09	1	0,00	0,05	0,05	0,08	-0,06	0,07	-0,07	0,04	-0,08	0,15	0,03	0,13	0,04	0,00	0,05	-0,03	0,04	-0,03	0,06
X207	0,01	-0,01	0,02	0,08	0,07	-0,03	-0,01	-0,11	0,04	0,10	0,00	-0,05	0,19	-0,01	0,10	0,06	0,00	1	-0,02	-0,01	-0,01	-0,06	0,01	0,06	0,05	0,10	0,13	0,01	0,03	-0,14	0,05	0,02	0,04	-0,02	0,00	0,13
X301	0,68	-0,04	0,80	0,63	-0,10	-0,07	-0,01	0,07	0,17	0,06	-0,01	-0,06	-0,06	-0,09	-0,64	-0,80	0,05	-0,02	1	0,75	0,68	0,48	0,79	-0,64	-0,01	0,08	-0,37	0,05	0,00	0,41	0,42	-0,31	-0,47	-0,19	0,44	0,02
X302	0,70	-0,02	0,80	0,68	-0,14	-0,06	-0,07	-0,02	0,08	0,00	0,01	0,07	-0,18	-0,09	-0,67	-0,84	0,05	-0,01	0,75	1	0,56	0,72	0,68	-0,63	-0,07	-0,03	-0,29	0,58	0,06	-0,19	-0,17	-0,44	-0,74	-0,39	0,43	-0,11
X303	0,51	-0,02	0,62	0,53	-0,11	-0,06	-0,03	-0,02	0,07	0,04	0,02	0,07	-0,06	-0,23	-0,45	-0,65	0,08	-0,01	0,68	0,56	1	0,24	0,67	-0,41	0,05	0,00	-0,17	0,06	-0,07	0,15	0,06	-0,15	-0,33	0,49	0,33	-0,10
X304	0,51	-0,03	0,55	0,44	-0,13	-0,03	-0,02	0,03	-0,09	-0,10	0,02	0,01	-0,08	0,03	-0,47	-0,57	-0,06	-0,06	0,48	0,72	0,24	1	0,25	-0,43	-0,03	0,06	-0,46	0,20	0,08	-0,15	-0,12	-0,56	-0,75	-0,45	0,45	-0,09
X305	0,60	-0,11	0,74	0,55	-0,15	-0,05	0,06	0,04	0,20	0,14	0,04	0,03	-0,12	-0,16	-0,59	-0,70	0,07	0,01	0,79	0,68	0,67	0,25	1	-0,55	-0,07	-0,04	-0,14	0,19	0,03	0,14	0,14	0,17	-0,15	-0,08	0,27	-0,05
X306	-0,58	-0,11	-0,54	-0,49	-0,11	0,18	-0,01	0,01	-0,12	-0,10	0,07	0,22	-0,12	-0,10	0,95	0,62	-0,07	0,06	-0,64	-0,63	-0,41	-0,43	-0,55	1	0,01	-0,02	0,23	-0,23	-0,01	-0,07	-0,15	0,28	0,40	0,28	-0,35	-0,21
X307	0,02	0,08	-0,05	-0,03	0,14	-0,19	-0,13	-0,10	0,11	0,00	-0,07	0,07	-0,06	-0,06	-0,02	0,04	0,04	0,05	-0,01	-0,07	0,05	-0,03	-0,07	0,01	1	-0,01	0,02	-0,10	-0,04	0,07	0,04	-0,06	0,04	0,12	0,01	-0,12
X308	0,02	-0,05	0,06	0,08	-0,06	0,05	0,05	0,13	-0,11	-0,05	-0,01	-0,01	0,05	0,04	-0,01	0,02	-0,08	0,10	0,08	-0,03	0,00	0,06	-0,04	-0,02	-0,01	1	0,04	-0,13	-0,05	0,15	0,17	-0,08	-0,02	-0,01	0,03	0,02
Y301	-0,32	-0,09	-0,28	-0,25	0,07	0,02	0,08	-0,03	0,07	0,05	0,02	0,03	-0,03	0,08	0,24	0,27	0,15	0,13	-0,37	-0,29	-0,17	-0,46	-0,14	0,23	0,02	0,04	1	-0,06	0,13	-0,17	-0,21	0,26	0,37	0,17	-0,33	-0,01
Y302	0,27	0,06	0,24	0,32	-0,03	0,01	-0,15	-0,11	-0,03	-0,02	-0,02	0,15	-0,21	-0,07	-0,29	-0,33	0,03	0,01	0,05	0,58	0,06	0,20	0,19	-0,23	-0,10	-0,13	-0,06	1	-0,09	-0,57	-0,55	-0,18	-0,45	-0,38	0,07	-0,16
Y303	-0,02	0,08	-0,02	-0,04	0,08	-0,12	0,19	-0,02	0,19	-0,15	0,07	-0,03	0,10	0,07	0,03	0,03	0,13	0,03	0,00	0,06	-0,07	0,08	0,03	-0,01	-0,04	-0,05	0,13	-0,09	1	0,03	-0,06	-0,01	0,08	-0,11	-0,08	0,12
Y304	0,03	0,02	0,05	-0,02	0,07	0,02	0,05	0,23	0,09	0,03	-0,03	-0,08	0,03	-0,05	-0,04	-0,05	0,04	-0,14	0,41	-0,19	0,15	-0,15	0,14	-0,07	0,07	0,15	-0,17	-0,57	0,03	1	0,52	0,02	0,20	0,16	0,04	0,08
Y305	0,08	-0,01	0,10	0,01	0,05	-0,06	0,10	0,00	0,15	0,07	0,00	-0,27	0,24	0,10	-0,07	-0,02	0,00	0,05	0,42	-0,17	0,06	-0,12	0,14	-0,15	0,04	0,17	-0,21	-0,55	-0,06	0,52	1	0,02	0,20	0,01	0,08	0,26
Y306	-0,30	-0,12	-0,27	-0,29	-0,09	0,08	0,17	0,05	0,08	0,10	-0,03	0,03	0,01	-0,04	0,27	0,39	0,05	0,02	-0,31	-0,44	-0,15	-0,56	0,17	0,28	-0,06	-0,08	0,26	-0,18	-0,01	0,02	0,02	1	0,57			

## Apêndice B – Rotinas utilizadas no Matlab

```
% PLS para múltiplos conjuntos X Y
% -> as tabelas devem ter nomes e distribuição como no formato do arquivo modelo e o arquivo deve estar no
formato .xls
% ARQUIVO: BASES2?.xls

%N2 = numero de linhas da porção treino (80% de 160)
N2=128;
% Entrar com o Número Do Estágio
NE=input('Qual o Número do Estágio?\n');
GRUPO = input('Qual o número do grupo?\n');
%Número de Conjuntos X Y
N=1;
% VL = Número de variáveis latentes
for VL = 1:20
    % VAREXCL= Número de variáveis excluídas
    VAREXCL=0;
    % arquivos originais lidos no Excel XX e YY com ajustes de células
    XX=cell(1,N);
    YY=cell(1,N);
    NX=cell(1,N);
    k=VAREXCL + 1;
    str1='Q';
    str2='X';
    str3='Y';
    str4=int2str(NE);
    str5=int2str(VAREXCL);
    str6=int2str(k);
    str7='A';
    str8=int2str(k+1);
    str9='F';
    if NE==1
        % estagio 1
        str10='K';
    elseif NE==2
        % estagio 2
        str10='R';
    elseif NE==3
        % estagio 3
        str10='Z';
    end
    str11='J';
    tab1=[str1 str4 str2 str5];
    tab2=[str1 str4 str3];
    tab3=[str1 str4 str2 str6];
    tab5=[str7 str8];
    tab6=[str7 str6 str10 str6];
    if GRUPO ==1
        % GRUPO a
        XX{1,N}=xlsread('BaseS2GA.xls',tab1);
        YY{1,N}=xlsread('BaseS2GA.xls',tab2);
        NX{1,N}=xlsread('BaseS2GA.xls','NomeX',tab6);
    elseif GRUPO ==2
        XX{1,N}=xlsread('BaseS2GB.xls',tab1);
        YY{1,N}=xlsread('BaseS2GB.xls',tab2);
        NX{1,N}=xlsread('BaseS2GB.xls','NomeX',tab6);
    elseif GRUPO ==3
        XX{1,N}=xlsread('BaseS2GC.xls',tab1);
        YY{1,N}=xlsread('BaseS2GC.xls',tab2);
        NX{1,N}=xlsread('BaseS2GC.xls','NomeX',tab6);
    end
end
```

```

elseif GRUPO ==4
    XX{1,N}=xlsread('BaseS2GD.xls',tab1);
    YY{1,N}=xlsread('BaseS2GD.xls',tab2);
    NX{1,N}=xlsread('BaseS2GD.xls','NomeX',tab6);
elseif GRUPO ==5
    XX{1,N}=xlsread('BaseS2GE.xls',tab1);
    YY{1,N}=xlsread('BaseS2GE.xls',tab2);
    NX{1,N}=xlsread('BaseS2GE.xls','NomeX',tab6);
end
% chama o PLS pra cada par e armazena os dados no workspace
X=cell(1,N);
X1=cell(1,N);
Y=cell(1,N);
Y1=cell(1,N);
SQR_tr=cell(1,N);
RMSE_tr=cell(1,N);
Residual_ts=cell(1,N);
SQR_ts=cell(1,N);
RMSE_ts=cell(1,N);
b=cell(1,N);
R2_Y=cell(1,N);
B_pls=cell(1,N);
Qy2=cell(1,N);
rank=cell(1,N);
VEXC=cell(1,N);

% Chama a rotina PLS_nipals para Validação Cruzada
[X{1,N},X1{1,N},Y{1,N},Y1{1,N},SQR_tr{1,N},RMSE_tr{1,N},Residual_ts{1,N},SQR_ts{1,N},RMSE_ts{1,N},
b{1,N},R2_Y{1,N},B_pls{1,N},Qy2{1,N}]...
    =PLS_nipals_vc(XX{1,N},YY{1,N},VL,N2);

%Calcular o rank das variáveis independentes
temp=abs(B_pls{1,N});
rank{1,N}=sum(temp,2);
%Identificar o menor rank, identificar a variável, guardar informações sobre a variável excluída
[nnx,nnp]=min(rank{1,N});
ve = NX{1,N}(1,nnp);
%Redefinir Nomes Variáveis X
tm = size(XX{1,N},2);
NXNOVO = [NX{1,N}(1,1:nnp-1),NX{1,N}(1,nnp+1:tm)];
XXNOVO = [XX{1,N}(:,1:nnp-1),XX{1,N}(:,nnp+1:tm)];
RMSE =
[RMSE_ts{1,N}(1),RMSE_ts{1,N}(2),RMSE_ts{1,N}(3),RMSE_ts{1,N}(4),RMSE_ts{1,N}(5),RMSE_ts{1,N}
(6),RMSE_ts{1,N}(7),RMSE_ts{1,N}(8),RMSE_ts{1,N}(9),RMSE_ts{1,N}(10)];
R2Y = [R2_Y{1,N}(VL)];
VEX = [VAREXCL, ve, nnx, nnp, Qy2, R2Y];
str8 = int2str(VL);
tab7=[str7 str8 str9 str8];
tab8=[str7 str8 str11 str8];
%Gravar resultados em arquivos Excel
if GRUPO ==1
    GRAVAN = xlswrite('BaseS2GA.xls',NXNOVO,'NomeX',tab5);
    GRAVAX = xlswrite('BaseS2GA.xls',XXNOVO,tab3);
    GRAVAE = xlswrite('BaseS2GA.xls',VEX,'RES',tab7);
    GRAVAR = xlswrite('BaseS2GA.xls',RMSE,'RMSE',tab8);
elseif GRUPO ==2
    GRAVAN = xlswrite('BaseS2GB.xls',NXNOVO,'NomeX',tab5);
    GRAVAX = xlswrite('BaseS2GB.xls',XXNOVO,tab3);
    GRAVAE = xlswrite('BaseS2GB.xls',VEX,'RES',tab7);
    GRAVAR = xlswrite('BaseS2GB.xls',RMSE,'RMSE',tab8);

```

```

elseif GRUPO ==3
    GRAVAN = xlswrite('BaseS2GC.xls',NXNOVO,'NomeX',tab5);
    GRAVAX = xlswrite('BaseS2GC.xls',XXNOVO,tab3);
    GRAVAE = xlswrite('BaseS2GC.xls',VEX,'RES',tab7);
    GRAVAR = xlswrite('BaseS2GC.xls',RMSE,'RMSE',tab8);
elseif GRUPO ==4
    GRAVAN = xlswrite('BaseS2GD.xls',NXNOVO,'NomeX',tab5);
    GRAVAX = xlswrite('BaseS2GD.xls',XXNOVO,tab3);
    GRAVAE = xlswrite('BaseS2GD.xls',VEX,'RES',tab7);
    GRAVAR = xlswrite('BaseS2GD.xls',RMSE,'RMSE',tab8);
elseif GRUPO ==5
    GRAVAN = xlswrite('BaseS2GE.xls',NXNOVO,'NomeX',tab5);
    GRAVAX = xlswrite('BaseS2GE.xls',XXNOVO,tab3);
    GRAVAE = xlswrite('BaseS2GE.xls',VEX,'RES',tab7);
    GRAVAR = xlswrite('BaseS2GE.xls',RMSE,'RMSE',tab8);
end
end;
disp('fim');

% FUNÇÃO PLS NIPALS – VALIDAÇÃO CRUZADA
function [X,X1,Y,Y1,SQR_tr,RMSE_tr,Residual_ts,SQR_ts,RMSE_ts,b,R2_Y,B_pls,Qy2] =
PLS_nipals_b(XX,YY,nfactor,N)
% disp('PLS_nipals_vc')
% Definição de tamanho para treino e teste
N=160;
dv=128;
% Definição das matrizes de treino e teste
X=XX(1:dv,:);
X1=XX(dv+1:N,:);
Y=YY(1:dv,:);
Y1=YY(dv+1:N,:);
X_ori=X;
Y_ori=Y;

if exist('nfactor')~=1;
    nfactor=rank(X);
end

M_X=mean(X);
M_Y=mean(Y);
S_X=std(X);
S_Y=std(Y);
X=zscore(X);
Y=zscore(Y);
[nn,np]=size(X);
[n,nq]=size(Y);

if nn~= n;
    error('Incompatible # of rows for X and Y');
end

% Precisão para convergência
epsilon=eps;

% Número de componentes retidos
% Conjuntos de Y
U=zeros(n,nfactor);
C=zeros(nq,nfactor);
% Conjuntos de X
T=zeros(n,nfactor);

```

```

P=zeros(np,nfactor);
W=zeros(np,nfactor);
b=zeros(1,nfactor);
R2_X=zeros(1,nfactor);
R2_Y=zeros(1,nfactor);
Xres=X;
Yres=Y;
SS_X=sum(sum(X.^2));
SS_Y=sum(sum(Y.^2));

for l=1:nfactor
    t=orth(Yres(:,l));
    t0=orth(rand(n,1)*10);
    u=t;
    nstep=0;
    maxstep=100;
    while ((t0-t)*(t0-t) > epsilon/2) && (nstep < maxstep);
        nstep=nstep+1;
        t0=t;
        w=orth(Xres'*u);
        t=orth(Xres*w);
        c=orth(Yres'*t);
        u=Yres*c;
    end;
    p=Xres'*t;
    b_l=((t'*t)^(-1))*(u'*t);
    b_l=u'*t;
    % Colocar em matrizes
    b(l)=b_l;
    P(:,l)=p;
    W(:,l)=w;
    T(:,l)=t;
    U(:,l)=u;
    C(:,l)=c;
    % Deflacionamento de X e Y
    Xres=Xres-t*p';
    Yres=Yres-(b(l)*(t*c'));
    R2_X(l)=(t'*t)*(p'*p)/SS_X;
    R2_Y(l)=(t'*t)*(b(l).^2)*(c'*c)/SS_Y;
end
X_rec=T*P';
Y_rec=T*diag(b)*C';
% PREDICAO de Y e X
Xori_rec=X_rec*diag(S_X)+(ones(n,1)*M_X);
Qx2=sqrt(sum(sum((X_ori-Xori_rec).^2)));
Yori_rec=Y_rec*diag(S_Y)+(ones(n,1)*M_Y);
Qy2=sqrt(sum(sum((Y_ori-Yori_rec).^2)));
Wstar=W*inv(P'*W);
B_pls=Wstar*diag(b)*C';
Bpls_star=[[-M_X;eye(np,np)]*diag(S_X.^(-1))*B_pls*diag(S_Y)];
Bpls_star(1,:)=Bpls_star(1,:)+M_Y;
Residual_tr=Y_ori-Yori_rec;
SQR_tr = sum(Residual_tr.^2);
RMSE_tr = sqrt(SQR_tr/nn);

% Testando a predição com a porção teste
X1_red=zscore(X1);
M_Y1=mean(Y1);
S_Y1=std(Y1);
[n11,m11]=size(X1_red);

```

```
y_hat_red=X1_red*B_pls;  
y_hat=y_hat_red*diag(S_Y1)+(ones(n11,1)*M_Y1);  
Residual_ts=Y1-y_hat;  
SQR_ts = sum(Residual_ts.^2);  
if n11 == 0  
    RMSE_ts = sqrt(SQR_ts/nn);  
elseif n11 > 0  
    RMSE_ts = sqrt(SQR_ts/n11);  
end
```

## Apêndice C - Limites de Especificação, de Controle e Índices de Capacidade do Processo

Variável	ESPECIFICAÇÃO			ESTATÍSTICAS			CONTROLE			CAPACIDADE	
	LIE	VN	LSE	Mínimo	Média	Máximo	DP	LIC	LSC	CP	CPk
<i>y</i> <sub>101</sub>	15500	16750	18000	12401	16700	21786	1761	11418	21982	0,24	0,23
<i>y</i> <sub>102</sub>	1,40	1,50	1,60	1,28	1,55	1,87	0,11	1,21	1,89	0,29	0,14
<i>y</i> <sub>103</sub>	3	12	21	3,91	14,16	42,05	6,58	-5,58	33,91	0,46	0,35
<i>y</i> <sub>104</sub>	5860	7100	8330	1707	5125	15018	2527	-2456	12705	0,16	-0,10
<i>y</i> <sub>105</sub>	2,50	2,90	3,30	2,264	2,922	3,684	0,274	2,099	3,744	0,49	0,46
<i>y</i> <sub>106</sub>	100	130	160	74,12	128,10	158,78	16,70	77,99	178,22	0,60	0,56
<i>y</i> <sub>201</sub>	13000	14050	15100	10340	13989	18255	1504	9477	18501	0,23	0,22
<i>y</i> <sub>202</sub>	2	2,15	2,3	1,84	2,15	2,54	0,14	1,74	2,56	0,37	0,37
<i>y</i> <sub>203</sub>	0	7	14	0,00	7,06	25,83	5,27	-8,74	22,87	0,44	0,44
<i>y</i> <sub>204</sub>	0	11	22	0,00	11,55	60,01	12,71	-26,58	49,68	0,29	0,27
<i>y</i> <sub>205</sub>	4	6	8	3,5	6,9	10,2	1,1	3,5	10,3	0,58	0,31
<i>y</i> <sub>206</sub>	1000	1300	1600	1024	1312	1682	138	898	1727	0,72	0,69
<i>y</i> <sub>301</sub>	4,5	5,65	6,8	4,64	5,62	6,76	0,44	4,31	6,94	0,88	0,86
<i>y</i> <sub>302</sub>	22,5	27	31,5	22,75	26,81	31,58	1,61	21,97	31,64	0,93	0,89
<i>y</i> <sub>303</sub>	2,7	3,4	4,1	2,68	3,41	4,14	0,30	2,51	4,30	0,79	0,78
<i>y</i> <sub>304</sub>	13,2	15,5	17,8	13,17	15,48	17,43	0,80	13,06	17,89	0,95	0,94
<i>y</i> <sub>305</sub>	9,5	12	14,5	9,82	11,98	14,46	0,83	9,50	14,47	1,01	1,00
<i>y</i> <sub>306</sub>	4,5	5,6	6,7	4,83	5,64	6,70	0,37	4,52	6,76	0,98	0,95
<i>y</i> <sub>307</sub>	4	5	6	4,25	4,97	6,17	0,38	3,81	6,12	0,87	0,84
<i>y</i> <sub>308</sub>	4	5	6	3,84	4,85	5,91	0,42	3,59	6,10	0,80	0,67
<i>y</i> <sub>309</sub>	2,3	2,85	3,4	2,47	2,91	3,43	0,19	2,35	3,47	0,99	0,88
<i>y</i> <sub>310</sub>	4	7	10	4,5	7,5	11,0	1,2	4,0	11,0	0,85	0,71





## Estágio 2

	X201	X202	X203	X204	X205	X206	X207	X208	X209	X210	X211	X212	X213	X214	X215	X216	X217	X218	X219	X220	X221	X222	X223	X224	X225	X226	X227	X228	X229	X230	X231	X232	X233	X234	X235	X236	X237	X238	X239	Y201	Y202	Y203	Y204	Y205	Y206	Y207	Y208	
X101	0,17	-0,18	0,03	0,15	-0,27	-0,11	-0,13	-0,13	-0,13	-0,13	-0,24	-0,02	-0,20	-0,54	-0,22	-0,09	-0,21	-0,27	0,20	0,06	0,33	-0,21	0,35	-0,11	-0,23	-0,22	0,03	0,01	0,16	0,10	0,00	0,26	0,24	0,15	0,11	0,23	0,28	0,28	0,31	0,08	0,18	-0,18	0,25	-0,02	0,04	-0,16	0,47	
X102	0,23	-0,14	0,10	0,15	-0,49	0,15	0,14	0,13	0,15	0,15	-0,22	0,08	-0,03	-0,30	0,28	0,03	-0,06	0,07	-0,22	0,03	0,28	-0,05	0,41	-0,01	-0,03	-0,04	0,17	-0,07	0,16	-0,01	0,00	0,22	0,22	0,13	-0,14	0,12	0,20	0,17	0,20	0,07	0,30	0,06	0,68	0,05	0,11	-0,14	0,39	
X103	0,07	-0,07	0,09	-0,02	-0,27	-0,10	-0,10	-0,06	-0,12	-0,12	0,13	0,41	-0,20	-0,13	-0,11	-0,33	-0,28	-0,03	-0,29	0,27	0,36	-0,15	0,39	0,23	-0,18	-0,17	0,26	-0,17	-0,07	-0,25	-0,25	0,13	0,06	-0,02	-0,36	-0,17	0,18	0,23	0,23	-0,21	0,19	-0,11	0,10	-0,13	-0,22	-0,14	0,28	
X104	0,32	0,10	0,03	0,16	-0,22	0,09	0,09	0,08	0,07	0,08	-0,24	-0,28	0,15	-0,52	0,26	0,15	0,05	-0,05	-0,19	-0,01	0,15	0,25	0,24	0,01	0,15	0,15	0,04	0,10	0,20	-0,21	-0,12	0,01	0,04	-0,03	0,12	0,04	0,01	-0,03	-0,01	-0,04	0,22	-0,01	0,56	0,02	0,20	0,01	0,22	
X105	-0,11	0,41	-0,16	-0,13	0,20	0,13	0,14	0,13	0,16	0,16	-0,11	-0,07	0,22	0,72	0,52	0,19	0,44	0,68	-0,30	-0,10	-0,02	0,36	-0,16	-0,05	0,24	0,21	-0,29	0,01	-0,30	-0,33	-0,10	-0,31	-0,28	-0,17	0,02	-0,22	-0,38	-0,39	-0,41	0,09	-0,06	0,08	0,08	-0,03	0,02	0,18	-0,35	
X106	-0,10	0,15	0,06	-0,10	-0,25	0,47	0,46	0,47	0,45	0,43	0,15	0,10	-0,03	-0,07	-0,20	-0,22	0,11	0,10	0,19	-0,39	-0,04	-0,23	-0,25	-0,20	-0,07	-0,07	-0,17	0,08	0,13	0,10	0,07	0,05	0,05	0,10	0,14	0,24	0,06	-0,03	0,00	0,21	0,09	0,43	0,17	0,31	0,31	-0,63	0,14	
X107	-0,09	-0,01	0,06	-0,21	0,29	0,07	0,07	0,02	0,01	0,01	-0,22	-0,36	1,00	-0,18	0,06	-0,26	0,77	0,36	-0,34	-0,08	-0,02	0,46	-0,06	0,08	0,99	0,98	-0,11	0,13	-0,04	-0,11	0,07	-0,20	-0,14	-0,16	0,21	-0,45	-0,82	-0,82	-0,84	-0,01	-0,07	0,03	0,13	0,09	0,00	0,55	-0,53	
X108	-0,02	-0,11	0,17	-0,11	-0,12	0,59	0,59	0,55	0,58	0,58	0,01	-0,30	0,52	-0,18	0,19	0,03	0,41	0,29	-0,12	-0,58	0,15	0,40	-0,06	0,06	0,49	0,48	-0,22	0,29	0,17	0,05	0,14	-0,12	-0,08	-0,09	0,33	-0,03	-0,40	-0,48	-0,46	0,01	0,13	0,47	0,40	0,49	0,46	0,22	-0,03	
X109	-0,14	-0,27	0,27	-0,20	0,11	0,49	0,49	0,50	0,50	0,50	0,32	-0,22	0,32	-0,04	0,30	0,24	0,49	0,18	0,06	-0,68	0,12	0,24	-0,16	-0,08	0,31	0,31	-0,46	0,41	0,04	0,07	0,17	-0,09	-0,06	-0,04	0,46	0,01	-0,42	-0,50	-0,47	0,03	0,04	0,43	0,04	0,63	0,42	0,17	-0,13	
X110	-0,11	-0,14	0,27	-0,17	0,03	0,39	0,39	0,39	0,38	0,38	0,30	-0,13	0,23	-0,14	0,11	0,06	0,36	0,07	-0,02	-0,49	-0,02	0,22	-0,08	0,02	0,19	0,19	-0,29	0,36	0,10	0,15	0,13	-0,18	-0,16	-0,18	0,21	-0,15	-0,34	-0,38	-0,37	0,03	0,02	0,31	-0,05	0,51	0,26	0,12	-0,12	
X111	-0,09	0,18	0,18	-0,21	0,01	0,21	0,22	0,26	0,23	0,23	0,23	-0,05	0,02	-0,11	0,09	0,14	0,09	-0,04	0,16	-0,35	-0,04	-0,04	-0,17	-0,17	-0,01	-0,01	-0,14	0,11	0,12	0,10	0,11	-0,01	0,03	0,07	0,27	0,21	0,02	-0,03	0,00	0,15	0,05	0,25	0,39	0,34	0,17	-0,20	0,02	
X112	-0,10	0,17	0,18	-0,21	0,01	0,18	0,19	0,23	0,19	0,19	0,23	-0,02	0,03	-0,09	0,09	0,12	0,10	-0,03	0,13	-0,31	-0,04	-0,04	-0,15	-0,16	0,00	0,00	-0,13	0,08	0,09	0,09	0,10	-0,02	0,02	0,06	0,24	0,19	0,00	-0,03	-0,01	0,15	0,05	0,23	0,40	0,31	0,14	-0,18	0,00	
X113	-0,02	0,15	-0,10	-0,10	-0,15	0,02	0,01	0,00	0,01	0,01	-0,10	-0,17	0,12	-0,05	0,08	0,02	0,14	0,12	-0,02	-0,02	0,01	0,00	-0,01	-0,11	0,12	0,12	-0,08	0,15	0,14	0,10	0,11	-0,10	-0,06	0,00	0,19	0,10	-0,06	-0,10	-0,07	0,16	0,08	-0,02	0,56	-0,05	0,03	-0,07	-0,07	
X114	-0,02	0,01	0,00	-0,04	-0,12	0,32	0,32	0,29	0,31	0,32	-0,03	-0,15	0,32	-0,03	-0,02	-0,05	0,21	0,20	-0,04	-0,24	-0,09	0,27	-0,17	-0,03	0,30	0,29	-0,14	0,07	0,07	0,06	0,14	-0,16	-0,12	-0,07	0,18	0,03	-0,22	-0,26	-0,25	0,09	0,06	0,27	0,62	0,15	0,26	0,09	-0,02	
X115	0,11	0,19	0,20	0,10	0,10	0,35	0,35	0,32	0,33	0,33	-0,02	-0,31	0,30	-0,37	-0,15	-0,11	0,06	-0,34	0,10	-0,25	0,03	0,12	0,00	-0,09	0,25	0,21	-0,03	-0,12	0,11	0,04	0,05	-0,13	-0,07	-0,04	0,07	-0,04	-0,04	-0,05	-0,07	0,06	-0,01	0,28	0,18	0,26	0,30	0,13	-0,05	
X116	0,21	0,27	0,05	0,13	0,19	0,35	0,35	0,29	0,36	0,36	-0,05	-0,62	-0,25	-0,35	0,00	0,19	0,12	-0,14	0,35	-0,34	0,09	0,20	-0,12	-0,13	0,23	0,25	-0,17	0,42	0,39	0,08	0,10	-0,12	-0,05	-0,04	0,45	0,09	-0,17	-0,24	-0,23	0,06	-0,04	0,23	0,10	0,25	0,46	0,13	0,01	
X117	0,09	0,16	-0,03	0,06	0,32	0,39	0,39	0,35	0,42	0,42	0,01	-0,47	0,08	-0,03	-0,03	0,24	0,09	0,02	0,43	-0,50	0,07	0,06	-0,23	-0,17	0,07	0,09	-0,28	0,41	0,25	-0,05	-0,01	-0,16	-0,11	-0,04	0,52	0,23	-0,13	-0,18	-0,18	0,09	-0,18	0,32	-0,03	0,28	0,59	0,49	0,09	0,10
X118	0,00	-0,02	0,15	0,11	-0,11	0,28	0,28	0,30	0,35	0,36	0,28	0,27	-0,58	0,32	-0,17	0,26	-0,50	-0,13	0,47	-0,25	-0,08	-0,26	-0,15	-0,18	-0,61	-0,61	-0,09	-0,12	-0,09	0,10	0,05	0,06	0,04	0,07	-0,03	0,29	0,39	0,40	0,40	0,15	0,00	0,23	0,03	0,25	0,24	-0,25	0,49	
X119	0,12	0,06	0,14	0,17	-0,16	0,55	0,55	0,57	0,61	0,62	0,23	0,27	-0,50	0,25	-0,08	0,25	-0,46	-0,05	0,36	-0,43	-0,04	-0,18	-0,09	-0,17	-0,53	-0,54	0,01	-0,14	-0,01	0,09	0,07	0,20	0,18	0,17	-0,10	0,27	0,42	0,41	0,40	0,15	0,03	0,47	0,19	0,36	0,42	-0,31	0,61	
X120	0,20	-0,01	0,13	0,19	-0,34	0,43	0,42	0,45	0,47	0,48	0,06	0,25	-0,63	-0,06	-0,09	0,23	-0,61	-0,26	0,36	-0,32	-0,02	-0,25	0,07	-0,13	-0,66	-0,66	0,12	-0,18	0,08	0,15	0,07	0,39	0,36	0,31	-0,08	0,42	0,65	0,64	0,64	0,10	0,12	0,36	0,33	0,27	0,36	-0,43	0,83	
X121	0,23	0,03	0,11	0,22	-0,28	0,51	0,50	0,52	0,55	0,56	0,05	0,21	-0,59	-0,05	-0,10	0,24	-0,57	-0,22	0,39	-0,40	-0,01	-0,23	0,05	-0,18	-0,62	-0,63	0,11	-0,16	0,11	0,13	0,07	0,38	0,35	0,31	-0,05	0,42	0,62	0,60	0,60	0,14	0,08	0,42	0,33	0,31	0,44	-0,41	0,83	
X122	0,23	0,04	0,08	0,22	-0,35	0,44	0,43	0,44	0,47	0,47	-0,02	0,18	-0,60	-0,21	-0,17	0,14	-0,64	-0,31	0,36	-0,30	0,04	-0,28	0,15	-0,22	-0,63	-0,63	0,20	-0,10	0,23	0,13	0,02	0,37	0,34	0,29	-0,04	0,46	0,69	0,67	0,67	0,15	0,11	0,36	0,32	0,23	0,39	-0,44	0,87	
X123	0,21	0,06	0,09	0,21	-0,36	0,44	0,43	0,45	0,47	0,48	0,00	0,18	-0,62	-0,19	-0,16	0,16	-0,65	-0,33	0,37	-0,32	0,04	-0,28	0,13	-0,22	-0,65	-0,65	0,18	-0,09	0,23	0,13	0,02	0,35	0,33	0,28	-0,02	0,48	0,70	0,67	0,68	0,15	0,11	0,37	0,30	0,24	0,39	-0,46	0,86	
X124	0,18	-0,05	0,13	0,04	0,00	0,23	0,24	0,31	0,23	0,22	0,50	0,22	-0,26	0,03	-0,06	-0,14	-0,12	0,04	0,14	-0,24	0,10	-0,15	-0,10	0,09	-0,27	-0,27	-0,07	-0,03	0,03	0,10	0,04	0,17	0,11	0,09	-0,05	0,15	0,17	0,13	0,15	-0,10	0,01	0,29	-0,09	0,28	0,21	-0,33	0,10	
X125	0,35	-0,01	0,10	0,18	-0,40	0,21	0,20	0,18	0,22	0,22	-0,40	0,03	-0,13	-0,27	0,31	0,19	-0,23	0,02	-0,18	-0,03	0,05	-0,07	0,34	-0,18	-0,14	-0,16	0,42	-0,24	0,25	0,05	0,02	0,35	0,35	0,23	-0,30	0,13	0,40	0,42	0,43	0,22	0,23	0,09	0,79	0,07	0,09	-0,18	0,51	
X126	0,28	-0,05	0,08	0,08	-0,35	0,12	0,11	0,09	0,12	0,13	-0,46	-0,01	-0,09	-0,22	0,41	0,23	-0,12	0,12	-0,23	0,00	-0,04	0,06	0,24	-0,19	-0,09	-0,11	0,38	-0,22	0,23	0,03	0,01	0,33	0,33	0,21	-0,28	0,12	0,34	0,36	0,37	0,22	0,20	0,00	0,79	0,03	-0,02	-0,19	0,39	
X127	-0,06	-0,35	-0,14	-0,01	0,01	-0,79	-0,79	-0,78	-0,77	-0,77	-0,13	0,30	-0,10	0,18	0,02	0,07	-0,03	-0,05	-0,15	0,69	-0,22	-0,19	0,19	-0,04	-0,06	-0,06	0,08	-0,32	-0,37	0,03	-0,02	0,01	-0,01	-0,05	-0,33	-0,22	-0,02	0,08	0,08	0,06	-0,04	-0,68	-0,16	-0,53	-0,67	0,20	-0,25	
X128	0,12	-0,06	-0,33	0,13	-0,29	-0,48	-0,49	-0,47	-0,48																																							



## Estágio 3

	X301	X302	X303	X304	X305	X306	X307	X308	X309	X310	X311	X312	X313	X314	X315	X316	X317	X318	X319	X320	X321	X322	X323	X324	X325	X326	X327	X328	X329	X330	X331	X332	X333	X334	X335	X336	X337	X338	X339	Y301	Y302	Y303	Y304	Y305	Y306	Y307	Y308
X101	0,28	0,28	0,28	0,21	0,36	0,47	0,07	0,35	0,36	0,35	0,36	0,37	0,19	0,00	0,27	0,32	-0,39	0,43	0,01	-0,08	-0,14	-0,02	0,07	0,03	0,01	0,49	0,05	0,02	-0,07	-0,13	-0,19	0,35	-0,10	0,29	0,26	0,29	-0,01	0,35	-0,11	-0,38	-0,03	0,46	-0,26	0,31	0,19	0,02	0,04
X102	0,61	0,23	0,22	0,06	0,31	0,35	0,07	0,19	0,19	0,19	0,20	0,18	0,37	0,10	-0,07	0,24	-0,32	0,15	0,31	0,30	-0,28	-0,15	0,19	-0,21	0,09	0,52	0,11	-0,13	0,16	-0,11	0,01	0,21	0,10	0,05	0,57	0,00	-0,62	0,05	-0,13	-0,02	-0,10	0,66	0,06	0,49	-0,06	-0,25	0,36
X103	0,04	0,19	0,20	-0,05	-0,14	-0,06	0,31	-0,05	-0,05	-0,05	-0,04	-0,03	-0,01	0,03	-0,02	0,04	0,01	0,15	0,19	0,15	-0,38	0,01	0,04	0,14	-0,01	0,25	0,11	0,13	-0,20	-0,24	0,20	0,00	-0,06	0,06	-0,17	0,00	0,17	0,26	-0,12	0,00	0,24	0,04	-0,08	0,04	-0,17	-0,16	0,27
X104	0,30	-0,01	-0,01	-0,02	0,63	0,68	-0,24	0,08	0,09	0,09	0,10	0,13	0,39	-0,09	0,19	0,03	-0,53	0,20	-0,09	-0,01	-0,29	0,02	0,28	-0,30	-0,22	0,31	-0,15	-0,08	0,02	0,02	-0,02	0,11	0,09	0,01	0,25	-0,14	-0,42	0,14	-0,04	-0,35	-0,17	0,34	-0,25	0,35	-0,02	0,06	0,06
X105	-0,07	-0,42	-0,41	-0,30	-0,20	-0,35	0,00	-0,50	-0,50	-0,50	-0,50	-0,53	0,30	0,48	0,17	-0,19	0,31	-0,37	0,12	0,07	0,07	-0,07	-0,43	-0,12	0,30	-0,47	-0,04	0,08	-0,03	0,49	0,11	-0,54	0,07	-0,35	-0,04	-0,34	-0,25	-0,61	-0,05	0,57	-0,12	-0,30	-0,06	-0,14	-0,34	0,32	0,00
X106	0,12	0,04	0,03	0,39	-0,10	-0,09	0,11	0,14	0,14	0,14	0,13	0,15	0,19	-0,09	-0,14	0,50	-0,03	0,18	0,28	0,08	-0,02	0,03	0,08	0,07	-0,06	0,08	-0,16	-0,42	-0,15	0,21	0,08	0,14	0,47	-0,35	0,47	0,50	-0,06	0,11	0,24	-0,01	-0,03	0,35	0,11	0,34	0,01	-0,22	-0,08
X107	-0,13	-0,78	-0,78	-0,20	0,32	0,17	-0,44	-0,19	-0,19	-0,19	-0,19	-0,15	0,42	-0,05	-0,08	-0,03	-0,13	-0,45	-0,04	0,12	-0,20	0,07	-0,22	-0,56	-0,52	-0,04	-0,18	0,07	0,19	0,42	0,11	-0,23	-0,01	-0,13	-0,09	-0,19	-0,08	-0,30	-0,16	0,18	-0,33	-0,07	0,01	-0,05	-0,03	0,20	0,05
X108	0,08	-0,40	-0,40	-0,10	0,37	0,30	-0,39	0,04	0,04	0,04	0,04	0,06	0,58	0,23	-0,11	0,28	-0,24	0,11	0,15	0,06	-0,06	0,16	-0,25	-0,34	-0,21	0,25	0,05	-0,51	-0,07	0,38	-0,08	-0,01	0,54	-0,47	0,07	-0,09	-0,18	-0,20	-0,21	0,00	-0,18	0,18	0,06	0,18	-0,02	-0,13	-0,17
X109	-0,15	-0,42	-0,42	-0,15	0,18	0,12	-0,27	-0,03	-0,02	-0,03	-0,04	-0,02	0,41	0,21	-0,12	0,22	-0,08	0,16	0,01	-0,08	0,09	0,09	-0,18	-0,18	-0,12	0,07	-0,08	-0,53	0,00	0,53	0,03	-0,06	0,50	-0,47	-0,13	-0,18	-0,03	-0,26	-0,09	-0,09	-0,12	-0,05	0,06	-0,10	0,02	-0,07	-0,35
X110	-0,21	-0,32	-0,32	-0,10	0,17	0,09	-0,24	0,01	0,02	0,01	0,01	0,04	0,32	0,30	-0,24	0,20	-0,11	0,11	0,23	0,16	-0,08	0,19	-0,19	0,00	-0,13	0,26	0,02	-0,35	-0,07	0,35	-0,05	-0,03	0,39	-0,35	-0,10	0,01	0,13	-0,06	0,02	-0,06	0,04	0,08	0,17	-0,12	0,03	-0,20	-0,09
X111	0,08	0,01	0,01	0,11	0,14	0,19	0,06	0,17	0,18	0,17	0,17	0,18	0,15	-0,10	0,09	0,06	-0,06	0,10	-0,03	-0,12	0,10	-0,02	0,29	-0,13	-0,10	0,01	-0,12	-0,27	-0,17	0,09	0,06	0,16	0,27	-0,15	0,15	0,13	-0,07	-0,03	0,17	-0,14	-0,14	0,09	-0,06	0,32	0,17	0,06	-0,23
X112	0,07	-0,01	-0,01	0,10	0,12	0,17	0,08	0,15	0,16	0,15	0,15	0,16	0,15	-0,10	0,07	0,03	-0,04	0,05	-0,02	-0,09	0,10	-0,04	0,29	-0,14	-0,10	0,00	-0,12	-0,23	-0,17	0,09	0,06	0,15	0,23	-0,13	0,15	0,12	-0,07	-0,05	0,16	-0,11	-0,14	0,08	-0,04	0,31	0,15	0,05	-0,19
X113	0,12	-0,04	-0,04	0,18	0,13	0,13	-0,07	0,14	0,15	0,14	0,14	0,15	0,24	0,13	0,07	0,01	-0,12	-0,01	-0,01	-0,05	0,02	0,04	0,07	-0,13	-0,10	0,01	-0,13	0,02	-0,07	0,09	-0,18	0,13	-0,01	0,09	0,27	0,09	-0,21	-0,09	-0,08	-0,01	-0,18	0,17	-0,05	0,50	0,07	0,09	-0,03
X114	0,02	-0,20	-0,20	-0,01	0,24	0,22	-0,20	0,03	0,03	0,03	0,03	0,04	0,24	0,01	-0,08	0,12	-0,07	-0,05	0,02	-0,05	0,11	0,08	-0,10	-0,24	-0,18	0,01	0,04	-0,24	-0,14	0,18	-0,18	0,00	0,27	-0,24	0,24	0,06	-0,21	-0,27	-0,12	0,08	-0,21	0,15	0,09	0,40	-0,03	-0,10	-0,12
X115	0,19	-0,05	-0,04	-0,05	0,27	0,21	-0,41	0,06	0,07	0,06	0,06	0,09	-0,12	-0,24	0,00	0,04	-0,36	0,09	-0,20	-0,09	-0,25	0,03	0,34	-0,20	-0,07	0,22	0,13	-0,30	0,06	0,02	-0,04	0,05	0,34	-0,27	0,08	0,07	-0,03	0,09	-0,01	-0,22	-0,22	0,17	-0,17	0,06	0,10	0,00	-0,19
X116	0,06	-0,19	-0,19	0,23	0,46	0,40	-0,59	0,26	0,25	0,29	0,25	0,05	0,10	0,22	-0,47	0,36	-0,14	-0,11	0,05	0,16	-0,03	-0,19	-0,13	0,09	0,05	-0,35	0,20	0,17	-0,15	0,21	0,31	-0,15	-0,03	-0,03	0,01	0,18	0,02	-0,34	-0,26	0,03	-0,11	0,18	0,33	0,06	-0,25		
X117	-0,08	-0,16	-0,14	0,21	0,21	0,20	-0,30	0,27	0,27	0,26	0,26	0,26	0,20	0,12	0,21	0,31	-0,14	0,45	-0,16	-0,25	0,25	0,20	-0,20	-0,06	0,02	-0,14	0,13	-0,44	0,05	0,21	-0,12	0,19	0,36	-0,21	-0,15	-0,08	0,10	-0,01	-0,07	-0,26	-0,21	-0,18	-0,14	0,08	0,40	0,11	-0,43
X118	-0,11	0,40	0,40	0,12	-0,21	-0,14	0,25	0,10	0,10	0,09	0,09	0,06	-0,30	-0,04	0,03	0,29	0,31	0,35	0,22	-0,01	0,47	-0,17	-0,08	0,18	0,46	-0,05	0,39	-0,36	-0,34	-0,12	-0,14	0,10	0,31	-0,21	-0,04	0,11	0,14	-0,11	0,10	0,00	-0,01	-0,06	0,18	0,08	0,02	-0,28	-0,12
X119	0,07	0,41	0,42	0,03	-0,13	-0,08	0,19	0,04	0,06	0,04	0,05	0,01	-0,22	-0,06	-0,01	0,39	0,19	0,40	0,32	0,08	0,31	-0,13	-0,03	0,12	0,49	0,03	0,40	-0,56	-0,35	-0,12	-0,04	0,07	0,58	-0,45	0,05	0,03	-0,03	-0,09	0,13	0,01	-0,03	0,04	0,17	0,13	-0,04	-0,36	-0,10
X120	0,30	0,62	0,63	0,08	0,01	0,16	0,27	0,18	0,19	0,18	0,19	0,15	-0,21	-0,15	0,12	0,38	-0,02	0,57	0,16	-0,02	0,13	-0,05	0,19	0,10	0,42	0,25	0,34	-0,50	-0,25	-0,30	-0,02	0,24	0,48	-0,26	0,23	0,16	-0,12	0,15	0,12	-0,28	0,00	0,28	-0,04	0,28	-0,01	-0,33	-0,11
X121	0,31	0,59	0,60	0,09	0,04	0,17	0,22	0,19	0,21	0,19	0,20	0,17	-0,19	-0,15	0,14	0,44	-0,05	0,59	0,18	-0,01	0,14	-0,05	0,15	0,09	0,45	0,24	0,35	-0,57	-0,26	-0,26	-0,03	0,24	0,55	-0,32	0,23	0,14	-0,14	0,13	0,12	-0,26	-0,04	0,28	-0,04	0,28	0,04	-0,31	-0,14
X122	0,41	0,65	0,66	0,20	0,10	0,26	0,20	0,33	0,34	0,33	0,34	0,31	-0,14	-0,08	0,16	0,50	-0,18	0,72	0,16	-0,03	0,01	0,07	0,13	0,15	0,37	0,34	0,36	-0,50	-0,19	-0,37	-0,04	0,37	0,47	-0,18	0,26	0,20	-0,10	0,30	0,10	-0,36	-0,04	0,36	-0,12	0,35	0,13	-0,32	-0,10
X123	0,40	0,66	0,67	0,20	0,09	0,25	0,21	0,33	0,34	0,33	0,34	0,31	-0,15	-0,07	0,16	0,49	-0,17	0,72	0,14	-0,06	0,03	0,08	0,13	0,17	0,37	0,31	0,36	-0,51	-0,20	-0,36	-0,05	0,36	0,48	-0,20	0,25	0,21	-0,09	0,30	0,11	-0,37	-0,02	0,34	-0,12	0,33	0,13	-0,31	-0,13
X124	0,06	0,12	0,12	0,11	-0,15	-0,09	0,31	0,05	0,05	0,05	0,04	0,05	-0,03	-0,14	-0,13	0,21	0,05	0,11	0,12	0,05	0,03	-0,03	0,17	0,18	0,02	0,14	-0,04	-0,36	-0,09	0,03	0,22	0,06	0,31	-0,25	0,17	0,20	-0,01	0,11	0,09	-0,02	0,34	0,18	0,13	0,02	0,05	-0,15	-0,05
X125	0,76	0,45	0,45	0,06	0,37	0,40	0,03	0,19	0,20	0,19	0,21	0,18	0,17	-0,05	0,16	0,17	-0,40	0,14	0,27	0,30	-0,34	-0,12	0,29	-0,26	0,29	0,50	0,02	-0,15	-0,04	-0,34	0,12	0,25	0,17	0,03	0,55	-0,07	-0,68	0,15	0,09	-0,05	-0,25	0,64	-0,13	0,58	-0,02	-0,08	0,32
X126	0,75	0,39	0,38	0,07	0,35	0,38	0,06	0,15	0,16	0,16	0,17	0,14	0,23	-0,02	0,25	0,07	-0,38	0,04	0,20	0,25	-0,36	-0,08	0,28	-0,29	0,26	0,44	-0,14	-0,06	-0,01	-0,28	0,21	0,21	0,07	0,09	0,58	-0,05	-0,70	0,13	0,16	-0,01	-0,27	0,62	-0,23	0,59	-0,03	0,07	0,28
X127	-0,15	0,07	0,04	-0,05	-0,23	-0,23	0,32	-0,11	-0,11	-0,11	-0,11	-0,14	-0,30	-0,11	-0,11	-0,47	0,35	-0,49	0,12	0,30	0,12	-0,50	0,02	-0,02	0,08	-0,07	-0,22	0,71	0,10	-0,23	-0,10	-0,08	-0,79	0,64	-0,08	-0,06	0,04	-0,10	-0,06	0,23	-0,06	-0,09	0,26	-0,11	-0,10	0,06	0,51
X128	0,08	0,27	0,27	0,06	-0,39	-0,34	0,51	-0,06	-0,07	-0,06	-0,05	-0,09																																			

	X301	X302	X303	X304	X305	X306	X307	X308	X309	X310	X311	X312	X313	X314	X315	X316	X317	X318	X319	X320	X321	X322	X323	X324	X325	X326	X327	X328	X329	X330	X331	X332	X333	X334	X335	X336	X337	X338	X339	Y301	Y302	Y303	Y304	Y305	Y306	Y307	Y308
X201	0,26	0,12	0,12	0,18	0,44	0,35	-0,32	0,13	0,12	0,13	0,14	0,17	0,14	-0,05	0,01	0,24	-0,56	0,25	0,24	0,32	-0,16	-0,09	0,18	-0,05	0,15	0,42	0,12	-0,24	0,15	-0,09	-0,29	0,12	0,20	-0,07	0,22	-0,16	-0,36	0,30	-0,24	-0,15	-0,02	0,38	0,07	0,26	0,21	-0,10	0,27
X202	0,10	0,01	0,02	0,02	0,12	0,04	-0,27	-0,12	-0,12	-0,12	-0,11	-0,09	-0,02	0,08	0,14	-0,01	-0,22	0,06	0,02	0,05	-0,10	0,06	0,08	0,06	0,16	-0,17	0,26	0,01	0,07	0,04	-0,11	-0,12	0,17	-0,21	0,05	-0,05	-0,09	0,05	0,16	0,00	0,06	-0,07	-0,11	-0,01	-0,05	0,11	0,05
X203	0,03	0,08	0,06	0,21	-0,09	-0,05	0,05	0,14	0,15	0,14	0,13	0,12	-0,13	-0,26	0,09	0,09	0,13	0,03	-0,03	-0,03	-0,06	0,13	0,24	-0,36	-0,06	0,19	0,20	-0,20	-0,07	-0,18	0,65	0,20	0,17	-0,04	-0,12	0,19	0,25	0,13	0,46	-0,04	-0,17	0,03	-0,14	0,15	0,05	0,06	-0,08
X204	0,10	0,21	0,21	-0,05	0,37	0,29	-0,37	0,07	0,05	0,07	0,07	0,10	-0,02	-0,04	-0,06	0,17	-0,46	0,21	0,14	0,16	0,01	-0,16	0,12	0,30	0,28	0,21	0,22	-0,13	0,12	-0,07	-0,59	0,04	0,10	-0,05	0,19	-0,16	-0,32	0,18	-0,25	-0,14	0,24	0,25	0,13	0,00	0,17	-0,16	0,18
X205	-0,45	-0,38	-0,38	-0,09	0,17	0,04	-0,29	-0,04	-0,05	-0,04	-0,05	-0,02	0,02	-0,20	0,29	0,09	0,03	-0,15	-0,18	-0,19	0,30	0,14	-0,17	0,03	0,05	-0,31	0,04	-0,09	0,08	0,23	0,11	-0,11	0,02	-0,10	-0,31	-0,21	0,17	-0,10	0,20	0,08	0,06	-0,37	-0,15	-0,27	0,43	0,45	-0,28
X206	0,28	0,04	0,05	0,01	0,13	0,07	-0,24	0,05	0,06	0,05	0,05	0,06	0,20	0,01	0,00	0,62	-0,24	0,42	0,23	0,08	-0,18	0,26	-0,04	-0,07	0,20	0,21	0,17	-0,92	-0,19	0,13	0,23	0,04	0,98	-0,82	0,12	-0,05	-0,18	0,07	0,14	-0,05	-0,04	0,19	-0,07	0,20	0,07	-0,20	-0,26
X207	0,26	0,03	0,04	0,00	0,14	0,08	-0,25	0,04	0,05	0,04	0,04	0,05	0,20	0,00	0,00	0,60	-0,23	0,41	0,23	0,07	-0,17	0,26	-0,03	-0,07	0,20	0,20	0,17	-0,92	-0,20	0,14	0,23	0,03	0,98	-0,83	0,11	-0,05	-0,17	0,06	0,15	-0,05	-0,04	0,18	-0,07	0,19	0,06	-0,19	-0,27
X208	0,23	0,05	0,06	-0,02	0,07	0,03	-0,17	0,01	0,03	0,01	0,02	0,02	0,15	-0,02	-0,03	0,59	-0,17	0,42	0,23	0,07	-0,15	0,23	-0,02	-0,04	0,20	0,18	0,17	-0,92	-0,24	0,14	0,25	0,02	1,00	-0,85	0,08	-0,04	-0,13	0,04	0,14	-0,05	0,00	0,14	-0,04	0,15	0,03	-0,23	-0,28
X209	0,28	0,08	0,09	0,00	0,13	0,07	-0,22	0,06	0,07	0,06	0,06	0,06	0,18	0,01	0,02	0,60	-0,20	0,43	0,23	0,06	-0,10	0,21	-0,04	-0,06	0,26	0,17	0,18	-0,93	-0,21	0,13	0,21	0,05	0,98	-0,81	0,09	-0,09	-0,19	0,02	0,14	-0,04	-0,06	0,14	-0,05	0,19	0,07	-0,20	-0,29
X210	0,28	0,08	0,09	0,00	0,13	0,08	-0,22	0,06	0,07	0,06	0,06	0,06	0,17	0,01	0,02	0,59	-0,20	0,43	0,22	0,05	-0,09	0,21	-0,03	-0,07	0,26	0,16	0,19	-0,93	-0,21	0,12	0,21	0,06	0,98	-0,81	0,09	-0,10	-0,20	0,02	0,13	-0,04	-0,06	0,14	-0,05	0,19	0,07	-0,20	-0,29
X211	-0,39	-0,01	-0,01	0,03	-0,42	-0,39	0,23	-0,02	-0,01	-0,03	-0,03	-0,02	-0,25	-0,10	-0,48	0,11	0,37	0,17	0,20	0,15	0,15	-0,13	0,01	0,25	-0,11	-0,09	0,06	-0,35	-0,21	0,00	0,17	-0,01	0,29	-0,28	-0,50	-0,05	0,47	0,05	0,02	0,02	0,34	-0,38	0,45	-0,25	-0,03	-0,45	0,01
X212	-0,13	0,24	0,24	-0,20	-0,72	-0,62	0,80	-0,26	-0,25	-0,26	-0,26	-0,29	-0,40	-0,10	-0,42	-0,07	0,68	-0,18	0,38	0,25	0,00	-0,26	-0,03	0,29	0,12	-0,09	0,20	0,13	-0,16	-0,25	0,28	-0,19	-0,05	-0,09	-0,07	0,19	0,24	-0,12	0,05	0,42	0,25	-0,10	0,39	-0,18	-0,39	-0,49	0,29
X213	-0,13	-0,77	-0,78	-0,20	0,32	0,17	-0,45	-0,19	-0,19	-0,19	-0,19	-0,15	0,42	-0,05	-0,08	-0,04	-0,13	-0,45	-0,04	0,12	-0,21	0,08	-0,21	-0,56	-0,51	-0,03	-0,18	0,07	0,19	0,41	0,11	-0,24	0,00	-0,14	-0,09	-0,19	-0,07	-0,29	-0,16	0,18	-0,33	-0,07	0,01	-0,05	-0,04	0,20	0,05
X214	-0,25	-0,11	-0,10	-0,31	-0,66	-0,70	0,45	-0,49	-0,49	-0,49	-0,50	-0,56	-0,19	0,28	0,04	-0,23	0,76	-0,40	0,08	-0,04	0,34	-0,16	-0,39	0,16	0,44	-0,62	0,08	0,13	-0,18	0,24	0,21	-0,50	-0,05	-0,24	-0,19	-0,19	0,04	-0,62	0,00	0,67	0,12	-0,49	0,10	-0,32	-0,37	0,13	-0,06
X215	0,25	-0,19	-0,18	-0,29	0,18	0,12	-0,13	-0,25	-0,25	-0,25	-0,24	-0,28	0,54	0,52	0,13	-0,37	-0,10	-0,16	0,03	0,05	-0,07	-0,05	-0,03	-0,26	0,17	-0,04	-0,25	0,03	0,22	0,26	0,15	-0,25	0,00	-0,11	0,05	-0,49	-0,51	-0,24	0,03	0,20	-0,20	0,00	-0,11	0,02	-0,27	0,21	0,02
X216	0,12	0,17	0,17	-0,12	0,35	0,35	-0,28	0,05	0,05	0,05	0,05	0,02	0,16	0,17	0,37	-0,21	-0,15	0,17	-0,16	-0,15	0,34	-0,27	0,08	-0,08	0,41	-0,12	-0,31	-0,15	-0,13	0,11	-0,20	0,04	0,05	0,01	-0,10	-0,40	-0,27	-0,10	0,19	-0,21	-0,16	-0,14	-0,17	0,01	0,05	0,29	-0,22
X217	-0,29	-0,87	-0,87	-0,37	0,18	0,05	-0,20	-0,43	-0,43	-0,43	-0,43	-0,41	0,54	0,22	0,03	-0,16	0,05	-0,49	-0,11	-0,06	-0,07	0,04	-0,31	-0,34	-0,35	-0,23	-0,40	0,15	0,18	0,76	0,06	-0,49	-0,05	-0,23	0,03	-0,12	-0,13	-0,51	-0,12	0,27	-0,24	-0,11	-0,08	-0,22	-0,23	0,33	-0,14
X218	-0,10	-0,49	-0,49	-0,17	-0,07	-0,23	0,00	-0,34	-0,35	-0,34	-0,34	-0,36	0,56	0,37	0,13	0,14	0,23	-0,36	0,35	0,26	-0,05	0,08	-0,49	-0,19	0,15	-0,14	-0,24	-0,08	-0,03	0,45	0,30	-0,39	0,14	-0,33	0,04	-0,21	-0,21	-0,34	0,03	0,55	-0,06	-0,06	-0,01	0,06	-0,14	0,27	0,14
X219	-0,08	0,22	0,21	0,47	0,12	0,21	0,02	0,38	0,38	0,38	0,37	0,37	-0,18	-0,39	0,39	0,33	-0,05	0,42	-0,23	-0,38	0,58	-0,08	0,21	-0,04	0,18	0,02	0,22	-0,32	0,07	-0,04	-0,05	0,36	0,13	0,06	0,09	0,43	0,29	0,26	0,27	-0,37	-0,12	0,06	-0,17	0,15	0,47	0,20	-0,44
X220	-0,07	0,07	0,05	-0,02	-0,19	-0,20	0,14	-0,15	-0,17	-0,15	-0,15	-0,15	-0,26	0,04	-0,18	-0,41	0,12	-0,38	0,04	0,24	-0,17	-0,20	0,02	0,10	-0,08	0,03	0,09	0,83	0,30	-0,28	-0,14	-0,13	-0,82	0,64	0,01	0,08	0,08	0,09	-0,12	0,18	0,14	0,03	0,17	-0,08	-0,23	-0,04	0,62
X221	0,25	0,08	0,11	0,01	-0,05	-0,07	-0,11	0,13	0,14	0,13	0,13	0,14	0,06	0,23	-0,18	0,12	-0,20	0,26	0,06	0,09	-0,09	-0,22	-0,07	0,13	0,07	0,04	0,15	-0,11	-0,04	0,02	-0,27	0,13	0,08	0,00	-0,25	-0,36	-0,06	0,01	-0,59	-0,03	0,18	-0,14	0,13	-0,08	0,06	-0,20	0,09
X222	-0,10	-0,64	-0,62	-0,68	0,42	0,31	-0,40	-0,60	-0,61	-0,60	-0,60	-0,57	0,43	0,38	0,08	-0,30	-0,18	-0,30	-0,17	-0,07	-0,11	0,05	-0,33	-0,22	-0,14	-0,06	-0,14	0,05	0,21	0,61	-0,15	-0,67	0,07	-0,41	-0,01	-0,17	-0,14	-0,34	-0,12	0,18	-0,06	-0,05	-0,13	-0,34	-0,48	0,21	-0,09
X223	0,35	0,22	0,24	-0,11	0,03	0,02	-0,04	0,06	0,07	0,06	0,07	0,07	-0,04	0,25	-0,38	-0,02	-0,30	0,17	0,29	0,43	-0,42	-0,26	-0,01	0,12	0,04	0,29	0,13	0,20	0,06	-0,25	-0,44	0,10	-0,15	0,21	-0,08	-0,38	-0,25	0,08	-0,63	-0,08	0,03	0,12	0,27	-0,01	-0,12	-0,43	0,52
X224	-0,29	-0,11	-0,12	-0,23	-0,05	-0,09	-0,10	-0,09	-0,11	-0,09	-0,09	-0,06	0,01	0,04	-0,26	-0,25	0,12	-0,22	0,08	0,18	-0,01	0,06	0,00	0,27	-0,11	-0,01	0,24	0,10	0,09	-0,06	0,02	-0,11	-0,08	0,01	-0,20	-0,06	0,16	0,09	0,06	0,12	0,78	-0,14	0,24	-0,25	-0,05	-0,15	0,23
X225	-0,14	-0,77	-0,77	-0,19	0,30	0,15	-0,43	-0,17	-0,17	-0,17	-0,17	-0,13	0,42	-0,04	-0,12	-0,06	-0,11	-0,46	-0,03	0,14	-0,21	0,08	-0,22	-0,54	-0,53	-0,06	-0,18	0,10	0,22	0,37	0,13	-0,21	-0,04	-0,09	-0,12	-0,25	-0,09	-0,29	-0,17	0,20	-0,30	-0,11	0,04	-0,05	-0,01	0,18	0,08
X226	-0,17	-0,76	-0,77	-0,12	0,30	0,16	-0,42	-0,10	-0,11	-0,11	-0,11	-0,07	0,44	-0,05	-0,14	-0,06	-0,11	-0,43	-0,02	0,15	-0,18	0,12	-0,20	-0,54	-0,57	-0,07	-0,17	0,10	0,24	0,33	0,14	-0,14	-0,06	-0,04	-0,15	-0,26	-0,08	-0,25	-0,15	0,16	-0,29	-0,12	0,07	-0,02	0,05	0,16	0,09
X227	0,41	0,56	0,55	0,26	-0,03	-0,04	-0,09	0,38	0,38	0,38	0,39	0,38	-0,30	-0,27	-0,16	0,19	-0,22	0,06	0,28	0,37	-0,34	0,22	0,25	0,02	0,08	0,26	0,23	-0,03	-0,11	-0,79	0,26	0,46	0,04	0,23	-0,04	-0,22	-0,18	0,43	0,20	0,01	0,06	0,13	0,10	0,33	0,29	-0,25	0,40
X228	-0,22	-0,24	-0,24	0,32	0,33	0,33	-0,33	0,42	0,41	0,42	0,42	0,45	0,56	0,25																																	



Apêndice F – Coeficientes de regressão PLS do estudo simulado

F.1 - Estágio 1

Var. retidas	Variável de resposta							
	y <sub>101</sub>	y <sub>102</sub>	y <sub>103</sub>	y <sub>104</sub>	y <sub>105</sub>	y <sub>106</sub>	y <sub>107</sub>	y <sub>108</sub>
x <sub>102</sub>	0,074	0,064	0,194	-0,019	0,036	-0,138	0,259	-0,247
x <sub>106</sub>	-0,047	-0,052	-0,138	-0,010	0,565	-0,061	0,515	-0,483
x <sub>113</sub>	0,126	0,104	-0,051	-0,010	0,042	-0,100	-0,063	0,093
x <sub>114</sub>	0,147	0,121	0,001	-0,011	0,036	-0,113	-0,033	0,061
x <sub>119</sub>	0,158	0,172	0,071	-0,005	-0,008	-0,111	0,022	-0,008
x <sub>122</sub>	0,143	0,190	0,076	-0,003	-0,014	-0,128	0,085	-0,071
x <sub>126</sub>	0,139	0,186	0,065	-0,002	0,004	-0,119	0,088	-0,075
x <sub>127</sub>	0,099	0,098	0,246	-0,009	-0,057	-0,083	0,154	-0,163
x <sub>128</sub>	0,011	0,028	-0,040	0,050	0,259	0,294	-0,037	-0,036
x <sub>130</sub>	0,020	0,051	-0,037	0,044	0,261	0,237	0,025	-0,086
x <sub>134</sub>	-0,044	0,062	-0,019	0,027	0,224	0,082	0,287	-0,318
x <sub>136</sub>	-0,033	0,068	0,050	0,005	-0,037	-0,048	0,196	-0,202
x <sub>138</sub>	-0,020	0,059	-0,065	-0,024	0,019	-0,233	0,269	-0,213

F.2 - Estágio 2

Var. retidas	Variável de resposta							
	y <sub>201</sub>	y <sub>202</sub>	y <sub>203</sub>	y <sub>204</sub>	y <sub>205</sub>	y <sub>206</sub>	y <sub>207</sub>	y <sub>208</sub>
x <sub>106</sub>	0,082	0,115	0,025	-0,027	-0,039	0,072	-0,502	-0,073
x <sub>108</sub>	-0,029	0,063	0,059	0,203	0,085	0,029	0,233	0,020
x <sub>109</sub>	-0,030	0,144	-0,031	-0,115	0,193	0,128	0,044	-0,105
x <sub>113</sub>	0,065	0,021	-0,022	0,426	-0,126	-0,013	-0,013	-0,023
x <sub>121</sub>	0,038	0,004	0,037	0,132	0,018	0,075	0,000	0,429
x <sub>125</sub>	0,075	0,058	0,006	0,354	0,039	0,002	-0,001	0,137
x <sub>126</sub>	0,078	0,042	-0,024	0,365	0,039	-0,027	-0,079	0,048
x <sub>127</sub>	0,108	-0,053	-0,080	0,052	-0,056	-0,127	0,082	-0,054
x <sub>131</sub>	0,014	0,032	0,490	0,059	0,057	0,051	0,005	-0,085
x <sub>134</sub>	-0,124	0,120	0,024	-0,109	0,053	0,093	-0,362	0,106
x <sub>203</sub>	-0,103	0,119	0,018	0,011	0,361	-0,141	0,094	-0,008
x <sub>204</sub>	-0,058	0,016	0,044	-0,014	-0,195	0,358	-0,070	0,108
x <sub>205</sub>	0,038	-0,311	0,046	-0,028	0,018	0,065	0,203	-0,074
x <sub>207</sub>	-0,007	-0,055	0,225	0,039	0,074	0,120	-0,007	0,045
x <sub>208</sub>	-0,020	-0,040	0,237	0,004	0,087	0,119	-0,023	0,045
x <sub>209</sub>	0,013	-0,051	0,226	0,043	0,093	0,119	0,014	0,059
x <sub>212</sub>	0,006	0,010	0,096	-0,035	-0,072	-0,143	-0,161	0,171
x <sub>219</sub>	0,070	0,011	-0,090	-0,056	0,088	0,156	-0,068	0,256
x <sub>220</sub>	-0,054	0,081	-0,148	0,018	-0,186	-0,083	-0,007	0,019
x <sub>222</sub>	-0,020	0,096	-0,003	0,084	-0,078	0,150	0,052	-0,050
x <sub>223</sub>	0,017	0,167	-0,003	-0,143	-0,096	0,064	0,183	0,069
x <sub>224</sub>	-0,622	0,043	-0,005	-0,018	-0,013	0,009	-0,009	0,031

## F.3 - Estágio 3

Var. retidas	Variável de resposta							
	$y_{301}$	$y_{302}$	$y_{303}$	$y_{304}$	$y_{305}$	$y_{306}$	$y_{307}$	$y_{308}$
$x_{102}$	0,013	0,075	0,145	0,007	0,075	-0,029	-0,020	0,056
$x_{106}$	0,035	0,028	0,074	0,062	0,089	-0,021	-0,087	-0,012
$x_{113}$	0,067	-0,017	0,009	-0,011	0,155	0,008	0,074	0,040
$x_{114}$	0,069	-0,061	0,028	0,072	0,127	-0,014	-0,066	-0,003
$x_{119}$	-0,043	-0,077	-0,007	0,084	0,029	-0,034	-0,163	-0,044
$x_{122}$	-0,148	-0,052	0,035	-0,003	0,022	-0,047	-0,129	-0,070
$x_{126}$	-0,010	-0,005	0,103	-0,039	0,121	-0,022	0,037	0,036
$x_{127}$	0,025	-0,031	0,016	0,069	0,012	0,005	0,061	0,166
$x_{128}$	0,072	-0,015	0,004	0,072	-0,011	-0,050	-0,071	0,135
$x_{130}$	0,099	0,021	-0,066	0,042	-0,035	0,073	0,032	-0,001
$x_{134}$	-0,120	0,051	-0,065	-0,037	0,004	-0,020	-0,051	-0,101
$x_{136}$	-0,035	0,110	0,064	-0,119	0,021	-0,025	0,047	-0,036
$x_{138}$	-0,038	-0,037	-0,060	-0,060	0,024	-0,064	0,091	0,056
$x_{204}$	-0,043	0,094	0,092	0,051	-0,018	0,075	-0,064	0,043
$x_{205}$	0,045	0,017	-0,029	-0,028	-0,046	0,198	0,123	-0,031
$x_{211}$	-0,021	0,109	-0,083	0,123	-0,068	-0,034	-0,145	0,006
$x_{212}$	0,149	0,072	-0,002	0,084	-0,051	-0,087	-0,130	0,047
$x_{214}$	0,219	0,049	-0,087	0,001	-0,001	-0,010	0,053	-0,008
$x_{218}$	0,193	0,029	0,014	0,000	0,081	0,067	0,084	0,054
$x_{219}$	-0,043	-0,064	0,037	-0,026	0,022	0,084	0,061	-0,060
$x_{221}$	0,098	0,074	-0,076	-0,084	-0,048	0,027	0,065	-0,025
$x_{223}$	-0,044	-0,065	-0,017	0,015	-0,044	-0,040	-0,072	0,074
$x_{224}$	-0,016	0,553	0,020	0,073	-0,038	0,007	-0,002	0,071
$x_{227}$	0,082	-0,038	-0,047	0,044	0,061	0,061	-0,056	0,075
$x_{301}$	0,093	-0,083	0,075	-0,073	0,085	-0,015	0,015	-0,032
$x_{304}$	0,021	-0,114	0,016	0,015	0,149	0,107	0,058	0,072
$x_{308}$	-0,044	0,033	0,004	0,038	0,079	0,163	-0,008	-0,015
$x_{309}$	-0,046	0,035	0,004	0,036	0,079	0,163	-0,007	-0,015
$x_{310}$	-0,051	0,030	0,003	0,038	0,080	0,160	-0,011	-0,015
$x_{311}$	-0,057	0,038	0,006	0,038	0,074	0,162	-0,011	-0,011
$x_{312}$	-0,008	-0,058	0,004	-0,327	0,044	0,045	0,384	-0,070
$x_{315}$	0,127	-0,009	0,051	0,147	0,078	0,028	-0,089	0,204
$x_{319}$	0,040	-0,001	0,043	0,174	0,078	0,004	-0,090	0,245
$x_{320}$	0,055	0,017	-0,056	0,128	0,022	0,073	-0,051	-0,061
$x_{321}$	-0,032	0,018	0,002	-0,076	-0,019	0,046	-0,028	-0,113
$x_{322}$	0,007	0,390	-0,021	-0,014	-0,098	0,039	-0,031	-0,042
$x_{324}$	0,044	0,086	0,020	-0,144	0,010	0,025	0,169	0,005
$x_{325}$	-0,032	-0,013	0,202	-0,050	0,023	0,006	0,020	0,101
$x_{326}$	0,156	0,018	0,059	0,036	-0,060	0,048	-0,081	0,005
$x_{327}$	0,036	-0,069	0,123	0,089	-0,024	0,013	-0,096	0,026
$x_{329}$	-0,026	-0,024	0,029	-0,070	-0,044	-0,062	0,082	-0,054
$x_{330}$	0,122	-0,056	-0,036	-0,045	0,039	-0,006	0,050	-0,013
$x_{331}$	0,039	0,045	0,226	0,024	0,109	-0,025	-0,041	0,008
$x_{335}$	0,019	0,069	0,174	-0,070	0,022	-0,051	0,071	0,033
$x_{336}$	-0,053	0,054	0,051	-0,068	0,023	0,047	0,051	0,068
$x_{338}$	0,013	0,075	0,145	0,007	0,075	-0,029	-0,020	0,056

## Apêndice G – Limites de Especificação, de Controle e Índices de Capacidade do Processo do estudo simulado

Variável	ESPECIFICAÇÃO			ESTATÍSTICAS			CONTROLE			CAPACIDADE	
	LIE	VN	LSE	Mínimo	Média	Máximo	DP	LIC	LSC	CP	CPk
<i>Y</i> <sub>101</sub>	0,80	3,00	5,20	-1,14	3,08	6,70	1,50	-1,41	7,56	0,49	0,47
<i>Y</i> <sub>102</sub>	0,75	1,75	2,75	-0,59	1,96	4,49	1,00	-1,02	4,95	0,33	0,26
<i>Y</i> <sub>103</sub>	1,00	4,75	8,50	0,77	4,96	8,84	1,65	0,01	9,91	0,76	0,71
<i>Y</i> <sub>104</sub>	4,00	10,00	16,00	0,32	9,35	16,46	3,05	0	19	0,66	0,58
<i>Y</i> <sub>105</sub>	-0,40	0,60	1,60	-2,32	0,54	2,70	0,71	-1,598	2,684	0,47	0,44
<i>Y</i> <sub>106</sub>	0,00	2,00	4,00	-0,71	1,89	5,14	1,10	-1,41	5,19	0,61	0,57
<i>Y</i> <sub>201</sub>	2,00	4,50	7,00	-0,82	4,13	8,10	1,62	-1	9	0,51	0,44
<i>Y</i> <sub>202</sub>	-4,00	-1,00	2,00	-6,20	-1,13	4,37	1,79	-6,52	4,25	0,56	0,53
<i>Y</i> <sub>203</sub>	-2,50	1,00	4,50	-2,54	0,89	4,61	1,44	-3,42	5,20	0,81	0,79
<i>Y</i> <sub>204</sub>	0,00	3,50	7,00	0,25	3,50	7,16	1,18	-0,03	7,03	0,99	0,99
<i>Y</i> <sub>205</sub>	6,00	7,50	9,00	3,24	7,01	9,99	0,97	4,1	9,9	0,52	0,35
<i>Y</i> <sub>206</sub>	2,00	4,00	6,00	0,99	3,82	6,15	0,91	1	7	0,73	0,67
<i>Y</i> <sub>301</sub>	3,50	5,50	7,50	0,41	5,10	8,09	1,24	1,37	8,83	0,54	0,43
<i>Y</i> <sub>302</sub>	9,00	12,00	15,00	7,23	11,95	16,00	1,36	7,87	16,03	0,74	0,72
<i>Y</i> <sub>303</sub>	-1,00	0,25	1,50	-1,70	0,46	3,32	0,71	-1,68	2,60	0,58	0,49
<i>Y</i> <sub>304</sub>	0,00	1,75	3,50	-1,14	1,85	5,74	1,20	-1,76	5,46	0,49	0,46
<i>Y</i> <sub>305</sub>	3,75	5,25	6,75	3,50	5,66	8,50	1,12	2,31	9,02	0,45	0,32
<i>Y</i> <sub>306</sub>	-2,00	0,00	2,00	-5,55	0,31	5,74	1,92	-5,45	6,07	0,35	0,29
<i>Y</i> <sub>307</sub>	-2,00	0,40	2,80	-4,82	-0,04	2,64	1,39	-4,22	4,13	0,57	0,47
<i>Y</i> <sub>308</sub>	7,20	7,80	8,40	6,20	8,04	10,55	0,58	6,29	9,79	0,34	0,21
<i>Y</i> <sub>309</sub>	1,40	3,00	4,60	0,00	2,69	4,79	0,85	0,13	5,25	0,62	0,50
<i>Y</i> <sub>310</sub>	9,50	11,00	12,50	5,02	10,21	14,11	1,77	4,9	15,5	0,28	0,13







